ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

J. Math. Pures Appl. 97 (2012) 318–390

JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES

www.elsevier.com/locate/matpur

Hamilton–Jacobi equations in space of measures associated with a system of conservation laws ☆

Jin Feng [a,b,*], Truyen Nguyen [a,b]

[a] *Mathematics Department, University of Kansas, Lawrence, KS 66045, USA*

[b] *Department of Theoretical and Applied Mathematics, University of Akron, Akron, OH 44325, USA*

Received 31 December 2010

Available online 28 November 2011

Abstract

We introduce a class of action integrals defined over probability measure-valued path space. We show that extremal point of such action exits and satisfies a type of compressible Euler equation in a weak sense. Moreover, we prove that both Cauchy and resolvent formulations of the associated Hamilton–Jacobi equations, in the space of probability measures, are well-posed.

Résumé

On introduit une classe d'intégrales d'action définies sur l'espace des chemins à valeurs mesures de probabilité. Dans ce contexte l'action minimale existe et donne une solution faible d'une équation d'Euler compressible. On montre que l'équation de Hamilton–Jacobi associée à la formulation variationnelle de l'équation d'Euler est bien posée dans le sens des solutions de viscosité.

MSC: primary 49L25; secondary 76N99, 60F10

Keywords: Compressible Euler equations; Mass transport; Hamilton–Jacobi equation; Viscosity solution in space of measures

1. Introduction

This article studies the well-posedness of a Hamilton–Jacobi partial differential equation in space of probability measures. Such equation is associated with a system of d-dimensional compressible Euler equations

☆ Jin Feng is supported in part by Army Research Office ARO W911NF-08-1-0064, and by National Science Foundation DMS-0806434; Truyen Nguyen is supported in part by National Science Foundation DMS-0901449 and by the University of Akron Faculty Research grant FRG-1716. We thank Giuseppe Savaré for useful discussions on (Gianazza et al., 2009 [17]) and Wilfrid Gangbo and the referee for helpful comments. Truyen Nguyen also acknowledges the hospitality of the Department of Mathematics, University of Kansas. We also thank Alessio Figalli and Andrzej Swiech for pointing out earlier errors which lead to the current version of Lemmas 2.10 and 2.15.

* Corresponding author at: Mathematics Department, University of Kansas, Lawrence, KS 66045, USA.
E-mail address: jfeng@math.ku.edu (J. Feng).

0021-7824/$ – see front matter
doi:10.1016/j.matpur.2011.11.004

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) + \nabla P(\rho) = -\rho \nabla(\phi + \Phi * \rho) - 2\nu^2 \rho \nabla\left(\dfrac{\Delta\sqrt{\rho}}{\sqrt{\rho}} - \dfrac{1}{4}\psi\right), \\ P(\rho) = \rho F'(\rho) - F(\rho), \end{cases} \tag{1.1}$$

where $\rho = \rho(t,x) : \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathbb{R}$, $u = u(t,x) : \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathbb{R}^d$ are unknown; $\nu > 0$ is a given constant and the functions $\psi, \phi, \Phi \in C^1(\mathbb{R})$, $F \in C^2(\mathbb{R}_+)$ are prescribed. Precise requirements on ψ, ϕ, Φ and F will be given in Condition 1.5. Notation $\operatorname{div}(\rho u \otimes u)$ should be understood as follows: this is a vector whose i-th component is $\operatorname{div}(\rho u u_i)$.

We prove two results. First, solution(s) to the above Euler equation will be derived as minimizer(s) of the following calculus of variation problem (Theorem 1.10):

$$\inf\left\{ g\big(\rho(t)\big) + \int_0^t L\big(\rho(s), \dot\rho(s)\big)\, ds \colon\ \rho(0) = \rho_0,\ \rho(\cdot) \in C\big([0,\infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big) \right\}. \tag{1.2}$$

Here $\mathcal{P}_2(\mathbb{R}^d)$ is the space of Borel probability measures on $\mathbb{R}^d$ with finite second moment, endowed with the Wasserstein-2 metric (e.g. [2]); g is a function from $\mathcal{P}_2(\mathbb{R}^d)$ to the extended reals $\mathbb{R} \cup \{+\infty\}$ which is prescribed carefully; $L(\rho, \dot\rho) = T(\rho, \dot\rho) - V(\rho)$, where

$$T(\rho, \dot\rho) = \frac{1}{2}\big\| \dot\rho - \nu\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big)\big\|^2_{-1,\rho},$$
$$V(\rho) = \int_{\mathbb{R}^d} \phi\, d\rho + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi(x-y)\rho(dx)\rho(dy) + \int_{\mathbb{R}^d} F(\rho)\, dx.$$

$\nu > 0$ is a positive parameter and the relation between the ψ in (1.1) and Ψ is that $\psi = |\nabla\Psi|^2 - 2\Delta\Psi$. The time derivative $\dot\rho$ for curves on $\mathcal{P}_2(\mathbb{R}^d)$ and the norm $\|\cdot\|_{-1,\rho}$ are defined in the standard way introduced in [2] (self-contained definitions will be given later in this article as well).

The second result is the characterization of the value function associated with this calculus of variation problem. If $f(t,\rho)$ denotes the minimum of (1.2), then f is the unique viscosity solution to the Hamilton–Jacobi equation

$$\begin{cases} \partial_t U(t,\rho) = H\big(\rho, \operatorname{grad} U(t,\rho)\big) & \text{in } (0,T) \times \mathcal{P}_2\big(\mathbb{R}^d\big), \\ U(t,\rho) = g(\rho) & \text{in } \mathcal{P}_2\big(\mathbb{R}^d\big), \end{cases}$$

where

$$H(\rho, n) := \sup_{m \in H_{-1,\rho}(\mathbb{R}^d)} \big(\langle n, m\rangle - L(\rho, m)\big),$$

and $\operatorname{grad} f(\rho)$ is the gradient of a map f defined on $\mathcal{P}_2(\mathbb{R}^d)$ [2] (a self-contained definition will appear later in this article). We also investigate the infinite horizon problem associated with (1.2), for which the Hamilton–Jacobi equation becomes stationary. See Theorems 1.13 and 1.14.

In Eq. (1.1), our point of view is to identify u with $\partial_t\rho$ through the transport equation, and then regard solution of (1.1) as describing a path $\rho(\cdot)$ with prescribed initial and terminal value of $\rho(0)$ and $\rho(T)$. This is in contrast with the usual partial differential equation formulation where $\rho(0), u(0)$ are given to start with, then we are asked to solve the equation to describe values of both $\rho(t)$ and $u(t)$ at later time $t > 0$. Therefore, there is a correspondence between $\rho(T)$ and $u(0)$, and more generally, between $\partial_t\rho(\cdot)$ and $u(\cdot)$. The arguments from (1.13) to (1.15) make it clear that we are essentially only considering $u(t,x)$ which is the closure of potential flows $\nabla_x\varphi(t,x)$ in some appropriate sense.

It is also useful to mention that (1.1) is not the usual nonlinear Schrödinger equation written under the Madelung transform – the forcing term has the "wrong" sign. After this article was completed, one of the authors attended a series of three talks given by P.L. Lions on mean-field games. The exact variational problem considered in this article also appeared in a broader context in the Lasry–Lions mean-field games theory [21]. Such a theory describes evolution of the $\rho(t)$ and a $\varphi(t,x)$ (in the $\nu = 0$ setting, φ is formally identified with $\partial_t\rho = -\operatorname{div}(\rho\nabla\varphi)$) in a system of equations with one written forward in time and the other backward in time. With the theory of Hamilton–Jacobi PDE in space of measures shown in this article, at least at a formal level, $\partial_t \rho$ (hence φ) and ρ are related by generating function $U(t,\rho)$ through relation (A.12) in Appendix A. Evolution of U is forward in time.

The rest of this introduction defines notations and setup of the problems rigorously.

First, at least on the surface, the term $\nu(\Delta\rho + \operatorname{div}(\rho\nabla\Psi))$ in the definition of $T(\rho,\dot\rho)$ appears odd in a continuum mechanic context. In the next subsection, we explain why it is natural using four related variational problems.

1.1. Three plus one variational problems

We start with a continuum mechanical system where (ρ, u) forms a closure. Kinetic energy for the system is

$$\tilde{T}(\rho,u) = \frac{1}{2}\int_{\mathbb{R}^d} \left|u(x)\right|^2 \rho(dx); \tag{1.3}$$

potential energy up to two particle interaction is

$$W(\rho) = \int_{\mathbb{R}^d} \phi\, d\rho + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi(x-y)\rho(dx)\,\rho(dy). \tag{1.4}$$

We also consider internal energy which, for $\rho(dx) = \rho(x)\,dx$ with a Lebesgue density, is defined as $\int_{\mathbb{R}^d} F(\rho(x))\,dx$. Let

$$V(\rho) = W(\rho) + \int_{\mathbb{R}^d} F\big(\rho(x)\big)\,dx, \tag{1.5}$$

and define Lagrangian and Hamiltonian respectively

$$\tilde{L}(\rho,u) = \tilde{T}(\rho,u) - V(\rho),$$
$$\tilde{H}(\rho,u) = \tilde{T}(\rho,u) + V(\rho),$$

and the corresponding action integral

$$\tilde{A}_T\big(\rho(\cdot),u(\cdot)\big) = \int_0^T \tilde{L}\big(\rho(t),u(t)\big)\,dt. \tag{1.6}$$

Let $\mathcal{P}(\mathbb{R}^d)$ denote the space of probability measures on $\mathbb{R}^d$, $\mathcal{P}_2(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$ the subspace with finite second moment, and $\mathcal{P}_2^r(\mathbb{R}^d) \subset \mathcal{P}_2(\mathbb{R}^d)$ the subspace with Lebesgue density. The above V is only defined for ρ with a Lebesgue density. Throughout this article, we will assume that V can also be extended to be defined for all $\rho \in \mathcal{P}(\mathbb{R}^d)$. For instance, when the following holds

$$\sup_{r>0}\frac{F(r)}{r} < \infty, \qquad \sup_{x\in\mathbb{R}^d}\Phi(x) + \sup_{x\in\mathbb{R}^d}\phi(x) < \infty, \tag{1.7}$$

then $V(\rho) \leqslant C < \infty$ for all ρ with Lebesgue density. We can extend the definition of V to those ρ without Lebesgue density in many ways while still keeping the property $\|V\vee 0\|_\infty < \infty$. It follows then $\tilde{L} : \mathcal{P}_2(\mathbb{R}^d)\times L^2_\rho(\mathbb{R}^d) \mapsto [-c,+\infty]$ with $c = \|V\vee 0\|_\infty$. $\tilde{A}$ becomes well defined. Moreover, we have $\tilde{H} : \mathcal{P}_2(\mathbb{R}^d)\times L^2_\rho(\mathbb{R}^d) \mapsto \mathbb{R}\cup\{+\infty\}$ at least when $V(\rho) > -\infty$.

Throughout this article, the space $\mathcal{P}_2(\mathbb{R}^d)$ is endowed with Wasserstein 2-metric d defined by

$$d^2(\rho,\gamma) = \inf\left\{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^2\pi(dx,dy)\colon \pi\in\Pi(\rho,\gamma)\right\},$$

where

$$\Pi(\rho,\gamma) := \left\{\pi(dx,dy)\in\mathcal{P}\big(\mathbb{R}^d\times\mathbb{R}^d\big)\colon \pi\big(dx,\mathbb{R}^d\big) = \rho(dx),\ \pi\big(\mathbb{R}^d,dy\big) = \gamma(dy)\right\}. \tag{1.8}$$

Then $(\mathcal{P}_2(\mathbb{R}^d), d)$ is a complete separable metric space (e.g. Proposition 7.1.5 of [2]). We recall also from p. 23 of [2] that a path $\sigma : (a, b) \mapsto \mathcal{P}_2(\mathbb{R}^d)$ is said to be p-absolutely continuous, $p \in [1, +\infty]$, if there exists a real-valued function $\beta \in L^p(a, b)$ such that

$$d\big(\sigma(t), \sigma(s)\big) \leqslant \int_s^t \beta(\tau)\, d\tau, \quad \forall a < s \leqslant t < b.$$

Let $AC^p(a, b; \mathcal{P}_2(\mathbb{R}^d))$ denote the space of all p-absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$. The case of $p = 1$ is called absolute continuous curves and denoted by $AC(a, b; \mathcal{P}_2(\mathbb{R}^d))$. For each $\rho(\cdot) \in AC(0, T; \mathcal{P}_2(\mathbb{R}^d))$, by the first part of Theorem 8.3.1 of [2], there exists vector field $u(t, \cdot) \in L^2_{\rho(t)}(\mathbb{R}^d)$ such that the continuity equation

$$\partial_t \rho + \operatorname{div}(\rho u) = 0 \tag{1.9}$$

holds in the sense of distributions. Moreover, if $\rho(\cdot) \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}^d))$, then by Theorem 1.1.2 and an estimate (8.3.13) in the proof of Theorem 8.3.1, both from [2],

$$\int_0^T \int_{\mathbb{R}^d} \big|u(t, x)\big|^2 \rho(t, dx)\, dt < \infty. \tag{1.10}$$

Given $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$, we denote

$$\Gamma(\rho_0, \rho_1) := \Gamma_T(\rho_0, \rho_1) := \big\{\sigma(\cdot) \in C\big([0, T]; \mathcal{P}_2\big(\mathbb{R}^d\big)\big)\colon\ \sigma(0) = \rho_0,\ \sigma(T) = \rho_1\big\}. \tag{1.11}$$

This is a closed subset of $C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$. With the above results in mind, we introduce our first variational problem

$$\inf\big\{\tilde{A}_T(\rho, u)\colon\ \rho(\cdot) \in AC^2\big(0, T; \mathcal{P}_2\big(\mathbb{R}^d\big)\big) \cap \Gamma(\rho_0, \rho_1), \partial_t \rho + \operatorname{div}(\rho u) = 0, u(t) \in L^2_{\rho(t)}\big(\mathbb{R}^d\big)\big\}. \tag{1.12}$$

At least formally, we can show that any minimizer of the above should satisfy (1.1) with $\nu = 0$. This is however not obvious and requires technical work.

We comment that informal derivation of Euler type equations out of variational problem has been known for quite some time (e.g. Chapter XI of Lanczos [20]). However, it seems that a rigorous derivation has not appeared until recently – see Gangbo, Nguyen and Tudorascu [15] for its derivation of a 1-D Euler–Poisson system. The result of this article is multi-dimensional and is based on arguments very different than that of [15]. A key is to introduce an extra regularization term which we motivate in detail below and in Appendix A. We also give well-posedness of Hamilton–Jacobi PDEs associated with the variational problem, whereas Gangbo, Nguyen and Tudorascu [16] gives only existence result.

The variational problem in (1.12) can be reduced. This requires a more careful study on the tangent space structure of $\mathcal{P}_2(\mathbb{R}^d)$, which was hinted by the work of Brenier [4], made explicit by Otto [22] and more extensively developed in Chapter 8 of Ambrosio, Gigli and Savaré [2]. Let

$$T_\rho \mathcal{P}_2\big(\mathbb{R}^d\big) = L^2_{\nabla,\rho}\big(\mathbb{R}^d\big) := \overline{\big\{\nabla\varphi\colon\ \varphi \in C_c^\infty\big(\mathbb{R}^d\big)\big\}}^{L^2_\rho(\mathbb{R}^d)} \subset L^2_\rho\big(\mathbb{R}^d\big). \tag{1.13}$$

Let $\rho(\cdot) \in AC(0, T; \mathcal{P}_2(\mathbb{R}^d))$ satisfy the continuity equation (1.9) with a vector field u satisfying (1.10). Let $v(t) = \Pi(u(t))$ be the projection of u onto $T_\rho \mathcal{P}_2(\mathbb{R}^d)$. Then we decompose

$$u = v + w, \quad v, w \in L^2_\rho\big(\mathbb{R}^d\big),\ v \in T_\rho \mathcal{P}_2\big(\mathbb{R}^d\big),\ w \in \big(T_\rho \mathcal{P}_2\big(\mathbb{R}^d\big)\big)^\perp.$$

It follows then (e.g. Proposition 8.4.3 and Remark 8.4.4 of [2])

$$\big(T_\rho \mathcal{P}_2\big(\mathbb{R}^d\big)\big)^\perp = \big\{w \in L^2_\rho\big(\mathbb{R}^d\big)\colon\ \operatorname{div}(\rho w) = 0\big\},$$

which implies that

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \quad v(t) \in T_{\rho(t)} \mathcal{P}_2\big(\mathbb{R}^d\big), \tag{1.14}$$

and that

$$\|u\|_{L^2_\rho(\mathbb{R}^d)} \geqslant \|v\|_{L^2_\rho(\mathbb{R}^d)}.$$

Putting everything together, $\tilde{T}(\rho(t), u(t)) \geqslant \tilde{T}(\rho(t), v(t))$ and $\tilde{A}_T(\rho, u) \geqslant \tilde{A}_T(\rho, v)$. Therefore, we arrive at a second variational problem which is equivalent to the first one:

$$\begin{aligned}
&\inf\bigl\{\tilde{A}_T(\rho, u)\colon\ \rho \in AC^2\bigl(0, T; \mathcal{P}_2\bigl(\mathbb{R}^d\bigr)\bigr) \cap \Gamma(\rho_0, \rho_1), \partial_t \rho + \operatorname{div}(\rho u) = 0, u(t) \in L^2_{\rho(t)}\bigl(\mathbb{R}^d\bigr)\bigr\} \\
&\quad = \inf\bigl\{\tilde{A}_T(\rho, v)\colon\ \rho \in AC^2\bigl(0, T; \mathcal{P}_2\bigl(\mathbb{R}^d\bigr)\bigr) \cap \Gamma(\rho_0, \rho_1), \partial_t \rho + \operatorname{div}(\rho v) = 0, v(t) \in T_{\rho(t)}\mathcal{P}_2\bigl(\mathbb{R}^d\bigr)\bigr\}.
\end{aligned} \tag{1.15}$$

In such infinite dimensional setting, $T_\rho \mathcal{P}_2(\mathbb{R}^d)$ can also be equivalently identified with dual spaces $H_{-1,\rho}(\mathbb{R}^d)$ and $H_{1,\rho}(\mathbb{R}^d)$ through isometry relations (e.g. Appendix D.5 of Feng and Kurtz [14]). Here and below, for $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, we define

$$\|\varphi\|^2_{1,\rho} = \int_{\mathbb{R}^d} |\nabla\varphi|^2 d\rho, \quad \varphi \in C_c^\infty\bigl(\mathbb{R}^d\bigr), \tag{1.16}$$

and

$$H_{1,\rho}\bigl(\mathbb{R}^d\bigr) = \text{completion of } C_c^\infty\bigl(\mathbb{R}^d\bigr) \text{ under } \|\cdot\|_{1,\rho}. \tag{1.17}$$

Let $m \in \mathcal{D}'(\mathbb{R}^d)$, the space of Schwartz distributions, we define

$$\|m\|^2_{-1,\rho} = \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \bigl\{2\langle m, \varphi\rangle - \|\varphi\|^2_{1,\rho}\bigr\}, \tag{1.18}$$

and

$$H_{-1,\rho}\bigl(\mathbb{R}^d\bigr) = \bigl\{m \in \mathcal{D}'\bigl(\mathbb{R}^d\bigr)\colon\ \|m\|_{-1,\rho} < \infty\bigr\}.$$

Throughout the rest of this article, although equivalent through transformation, we identify $T_\rho \mathcal{P}_2(\mathbb{R}^d)$ with $H_{-1,\rho}(\mathbb{R}^d)$ instead of $L^2_{\nabla,\rho}(\mathbb{R}^d)$ as earlier. Such identification was initially used by Otto [22] and has the advantage of being able to express many useful quantities in terms of Schwartz distribution, in the absence of further a priori estimates.

With a slight abuse of notation, we denote the Schwartz distributional derivative (in time) $\partial_t \rho = \dot{\rho}$. Note that in continuum mechanics, the dot notation usually denotes material derivative, which is not what intended here. Let $\rho(\cdot) \in AC(0, T; \mathcal{P}_2(\mathbb{R}^d))$ and v satisfy (1.14) and (1.10) (with u replaced by v). By Lemma D.34 of [14],

$$\bigl\|v(t)\bigr\|^2_{L^2_{\rho(t)}} = \bigl\|\dot{\rho}(t)\bigr\|^2_{-1,\rho(t)}.$$

Therefore

$$\tilde{A}_T(\rho, v) = \tilde{A}_T(\rho, \dot{\rho}) = \int_0^T \left(\frac{1}{2}\bigl\|\dot{\rho}(t)\bigr\|^2_{-1,\rho(t)} - V\bigl(\rho(t)\bigr)\right) dt. \tag{1.19}$$

The dependence of v is replaced by $\dot{\rho}$. We arrived at yet another equivalent way of writing the variational problem (1.12)

$$\begin{aligned}
&\inf\bigl\{\tilde{A}_T(\rho, u)\colon\ \rho(\cdot) \in AC\bigl(0, T; \mathcal{P}_2\bigl(\mathbb{R}^d\bigr)\bigr) \cap \Gamma(\rho_0, \rho_1), \partial_t \rho + \operatorname{div}(\rho u) = 0, u(t) \in L^2_{\rho(t)}\bigl(\mathbb{R}^d\bigr)\bigr\} \\
&\quad = \inf\left\{\int_0^T \tilde{L}\bigl(\rho(t), \dot{\rho}(t)\bigr)\, dt\colon\ \rho(\cdot) \in AC\bigl(0, T; \mathcal{P}_2\bigl(\mathbb{R}^d\bigr)\bigr) \cap \Gamma(\rho_0, \rho_1)\right\}.
\end{aligned}$$

While we can formally derive the extremal point of the above variational problem as solution to (1.1) with $\nu = 0$, there is no apparent way to make it rigorous. Indeed, showing existence of minimizer is even a nontrivial problem. Furthermore, it is even more out of reach for rigorously proving the uniqueness of Hamilton–Jacobi PDEs associated with such variational problem, which is another main result of this article. However, if we add a bit of "regularization" to our minimization problem in the particular way described below, we can prove useful theorems.

Let Lagrangian $L : \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{D}'(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{+\infty\}$ be defined by

$$L(\rho, \dot{\rho}) = T(\rho, \dot{\rho}) - V(\rho), \tag{1.20}$$

where

$$T(\rho, \dot{\rho}) = \frac{1}{2} \big\| \dot{\rho} - \nu\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big)\big\|_{-1,\rho}^2. \tag{1.21}$$

Then L is bounded from below since we assumed that V is bounded from above. We shall define a geometrically motivated notion of gradient for functions on $\mathcal{P}_2(\mathbb{R}^d)$ (see Definition 1.2), introduce an entropy function S in (1.29) and compute its gradient (1.30). It follows that if the "kinetic energy" functional

$$K_T\big[\rho(\cdot)\big] = \int_0^T T\big(\rho(t), \dot{\rho}(t)\big)\,dt \tag{1.22}$$

is finite, then for every $t > 0$, $S(\rho(t)) < \infty$ (in particular $\rho(t) \in \mathcal{P}_2^r(\mathbb{R}^d)$), and

$$\operatorname{grad} S\big(\rho(t)\big) = -\big(\Delta\rho(t) + \operatorname{div}\big(\rho(t)\nabla\Psi\big)\big),$$
$$T\big(\rho(t), \dot{\rho}(t)\big) = \frac{1}{2} \big\| \dot{\rho}(t) + \nu \operatorname{grad} S\big(\rho(t)\big)\big\|_{-1,\rho(t)}^2.$$

See Lemma 2.1. The observation that Fokker–Planck operator is gradient of entropy was first made by Jordan, Kinderlehrer and Otto [18].

We define an action integral $A_T : C([0,T]; \mathcal{P}_2(\mathbb{R}^d)) \mapsto \mathbb{R} \cup \{+\infty\}$ by

$$A_T\big[\rho(\cdot)\big] = \int_0^T L\big(\rho(s), \dot{\rho}(s)\big)\,ds. \tag{1.23}$$

Again, to make sure we do not end up with $\infty - \infty$, we simply assume (1.7) which is not optimal in terms of what we can do later, but simplifies the presentation to highlight on important matters at this stage. We will prove that minimizer of

$$\mathbb{D}(\rho_1, \rho_0; T) := \inf\big\{A_T\big[\rho(\cdot)\big]:\ \rho(\cdot) \in \Gamma(\rho_0, \rho_1)\big\} \tag{1.24}$$

exists (Lemma 3.1) and that, when ρ_0, ρ_1 satisfy a mild condition expressed in terms of entropy, any such minimizer is a Schwartz distributional solution to the system (1.1) (Theorem 1.10).

Remark 1.1. With $\nu > 0$, the extra perturbation in L introduces a preferred direction $-\nu \operatorname{grad} S(\rho)$ (i.e. entropy dissipation direction), for choices of $\dot{\rho}$, in order to minimize action. Such entropy dissipation property will produce a number of useful inequalities which help us to rigorously prove the main theorems in this article. See also Appendix A for a probabilistic origin of such regularization term.

In (1.33), we will introduce Fisher information functional $I \geqslant 0$. For $\rho(\cdot) \in AC^2(0,T; \mathcal{P}_2(\mathbb{R}^d))$ with $\int_0^T I(\rho)\,dt < \infty$, by the chain rule in 10.1.2 of [2],

$$\int_0^T \big\langle \operatorname{grad} S(\rho), \dot{\rho}\big\rangle_{-1,\rho}\,dt = S\big(\rho(T)\big) - S\big(\rho(0)\big),$$

consequently

$$A_T\big[\rho(\cdot)\big] = \int_0^T \left(\frac{1}{2}\|\dot{\rho}\|_{-1,\rho}^2 + \frac{\nu^2}{2} I(\rho) + \nu\big\langle \operatorname{grad} S(\rho), \dot{\rho}\big\rangle_{-1,\rho} - V(\rho)\right) dt$$
$$= \nu\big(S\big(\rho(T)\big) - S\big(\rho(0)\big)\big) + \int_0^T \left(\frac{1}{2}\|\dot{\rho}\|_{-1,\rho}^2 + \frac{\nu^2}{2} I(\rho) - V(\rho)\right) dt.$$

From this point on, we will take the convention that infimum of any function over an empty set is $+\infty$. Then, by a result in Lemma 3.2, with a mild condition on ρ_0, ρ_1, assuming $\|V \vee 0\| < \infty$, it follows that

$$\begin{aligned}\mathbb{D}(\rho_1, \rho_0; T) = \inf\left\{\int_0^T \left(\frac{1}{2}\|\dot{\rho}\|_{-1,\rho}^2 + \frac{\nu^2}{2} I(\rho) - V(\rho)\right) dt\colon\ \rho(\cdot) \in \Gamma(\rho_0, \rho_1)\right\}\\ + \nu\big(S(\rho_1) - S(\rho_0)\big).\end{aligned} \tag{1.25}$$

Moreover, the variational problem on the right hand side is well posed and attains the extremal point if $|V(\rho)| \leqslant \zeta(I(\rho))$ for some nondecreasing sub-linear function $\zeta \in C(\mathbb{R}_+)$ and if V is continuous on finite level sets of I. We will assume additional assumptions on F, Φ, ϕ in Condition 1.5 to ensure that this is the case (Lemma 2.18).

1.2. A special calculus on space of probability measures

Systematic accounts for modern mass transport theory can be found in, for instance, Ambrosio, Gigli and Savaré [2], Villani [24,25]. Below we selectively discuss and extend particular techniques which will be useful in this article.

Identifying $T_\rho \mathcal{P}_2(\mathbb{R}^d)$ with $H_{-1,\rho(t)}(\mathbb{R}^d)$, we introduce a compatible notion of gradient for functions on $\mathcal{P}_2(\mathbb{R}^d)$:

Definition 1.2 *(Gradient).* Let $f : \mathcal{P}_2(\mathbb{R}^d) \mapsto [-\infty, +\infty]$ and $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$. We say that gradient of f at ρ_0, denoted $\operatorname{grad} f(\rho_0)$, exists, if it can be identified as the unique element in $\mathcal{D}'(\mathbb{R}^d)$ such that for each $p \in C_c^\infty(\mathbb{R}^d)$, and each $\{\rho^p(t) \in \mathcal{P}_2(\mathbb{R}^d)\colon\ t \geqslant 0\}$ satisfying the continuity equation generated by p:

$$\partial_t \rho^p + \operatorname{div}\big(\rho^p \nabla p\big) = 0, \quad \rho^p(0) = \rho_0 \quad \text{in } \mathcal{D}'\big(\mathbb{R}^d\big), \tag{1.26}$$

we have

$$\lim_{t\to 0+} \frac{f(\rho^p(t)) - f(\rho^p(0))}{t} =: \big\langle \operatorname{grad} f(\rho_0), p\big\rangle.$$

Example 1.3. The relative entropy of $\rho \in \mathcal{P}(\mathbb{R}^d)$ with respect to $\gamma \in \mathcal{P}(\mathbb{R}^d)$ is defined as

$$\begin{aligned}S(\rho\|\gamma) &:= \begin{cases}\int_{\mathbb{R}^d} d\rho \log \frac{d\rho}{d\gamma} & \text{if } \frac{d\rho}{d\gamma} \in L^1_\gamma(\mathbb{R}^d)\\ +\infty & \text{otherwise}\end{cases}\\ &= \sup_{f\in C_b(\mathbb{R}^d)} \left\{\int_{\mathbb{R}^d} f\, d\rho - \log \int_{\mathbb{R}^d} e^f\, d\gamma\right\}.\end{aligned} \tag{1.27}$$

The variational representation in the second equality above can be found in, for instance, [11]. It follows then $S(\cdot\|\gamma) : \mathcal{P}(\mathbb{R}^d) \mapsto [0, +\infty]$ is well defined and lower semicontinuous with respect to the weak convergence of probability measure topology, also known as the narrow topology. Let

$$\mu^\Psi(dx) = Z^{-1} e^{-\Psi(x)}\, dx, \quad Z = \int_{\mathbb{R}^d} e^{-\Psi}\, dx. \tag{1.28}$$

Throughout this article, we denote

$$S(\rho) = S\big(\rho\|\mu^\Psi\big) = \int_{\mathbb{R}^d} \rho(x)\big(\log \rho(x) + \Psi(x)\big)\, dx + \log Z. \tag{1.29}$$

Assuming Ψ has super-quadratic growth at infinity (part of Condition 1.5), S has compact level sets in $\mathcal{P}_2(\mathbb{R}^d)$ (Lemma 9.39 of [14]). Provided $S(\rho) < \infty$, it follows (e.g. (9.98) in [14]) that

$$\operatorname{grad} S(\rho) = -\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big) \in \mathcal{D}'\big(\mathbb{R}^d\big). \tag{1.30}$$

Recall the definition of (1.4) and (1.5), under Condition 1.5,

$$\operatorname{grad} W(\rho) = -\operatorname{div}\big(\rho\nabla(\phi + \Phi * \rho)\big) \in H_{-1,\rho}\big(\mathbb{R}^d\big), \tag{1.31}$$

$$\operatorname{grad} V(\rho) = \operatorname{grad} W(\rho) - \Delta P(\rho) \in \mathcal{D}'\big(\mathbb{R}^d\big). \tag{1.32}$$

We introduce relative Fisher information

$$I(\rho\|\gamma) := \Big\| \operatorname*{grad}_{\rho} S(\rho\|\gamma)\Big\|^2_{-1,\rho}.$$

In particular, we denote

$$\begin{aligned} I(\rho) &:= \|\operatorname{grad} S(\rho)\|^2_{-1,\rho} = \int_{\mathbb{R}^d} \left|\nabla \log \frac{d\rho}{d\mu^{\Psi}}\right|^2 d\rho \\ &= \int_{\mathbb{R}^d} \frac{|\nabla\rho + \rho\nabla\Psi|^2}{\rho(x)}\,dx = \int_{\mathbb{R}^d} \frac{|\nabla\rho|^2}{\rho}\,dx + \int_{\mathbb{R}^d} \psi\,d\rho, \end{aligned} \tag{1.33}$$

where

$$\psi(x) := |\nabla\Psi|^2 - 2\Delta\Psi. \tag{1.34}$$

Equivalence of the first three expressions in (1.33), as well as some properties of I, are discussed in Appendix D.6 of [14]. In particular, $I : \mathcal{P}_2(\mathbb{R}^d) \mapsto [0, +\infty]$ is lower semicontinuous. Next, we outline a proof showing

$$I_1(\rho) := \int_{\mathbb{R}^d} \frac{|\nabla\rho + \rho\nabla\Psi|^2}{\rho(x)}\,dx = \int_{\mathbb{R}^d} \frac{|\nabla\rho|^2}{\rho}\,dx + \int_{\mathbb{R}^d} \psi\,d\rho := I_2(\rho).$$

Hence conclude the last equality in (1.33). First, if $\rho \in C^1_c(\mathbb{R}^d)$, then $I_1(\rho) < \infty$ (respectively $I_2(\rho) < \infty$) if and only if $\int_{\mathbb{R}^d} \frac{|\nabla\rho|^2}{\rho}\,dx < \infty$. In this case, we only need to show

$$\int_{\mathbb{R}^d} \frac{\nabla\rho}{\rho}\nabla\Psi\,d\rho = -\int_{\mathbb{R}^d} \Delta\Psi\,d\rho,$$

which follows from integration by parts. Secondly, by mollification and lower semicontinuity of I_1 and I_2 in the weak convergence of probability measure topology (i.e. the narrow topology), and by a convexity argument as in Lemma 8.1.10 in [2], $I_1(\rho) = I_2(\rho)$ for all ρ with bounded density and compact support. Finally, invoking approximation results in Lemma D.47 on p. 394 of [14], $I_1 = I_2$ for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$.

By Corollary 4.1 in [17] (taking $f = \frac{1}{2}\psi$ and $\mathcal{F}^f = \frac{1}{2}I$) and by an observation (1.56), at least formally,

$$\operatorname{grad} I(\rho) = -\operatorname{div}\left(\rho\nabla\left(-4\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} + \psi\right)\right), \quad \text{if } I(\rho) < \infty. \tag{1.35}$$

See also Example 11.1.10 in [2]. We will not use this particular result in a direct way, therefore not attempt to make it rigorous. However, knowing this expression helps to understand the last few terms in compressible Euler equation (1.1).

Example 1.4. Let $\varphi_j \in C^\infty_c(\mathbb{R}^d)$, $j = 1, \ldots, k$, and $f \in C^1(\mathbb{R}^k)$. With slight abuse of notation, we denote

$$f(\rho) := f\big(\langle\rho, \varphi_1\rangle, \ldots, \langle\rho, \varphi_k\rangle\big). \tag{1.36}$$

Then

$$\begin{aligned} \operatorname{grad} f(\rho) &= -\operatorname{div}\left(\rho\nabla\frac{\delta f}{\delta\rho}\right) \\ &= \sum_{j=1}^{k} \partial_j f\big(\langle\rho, \varphi_1\rangle, \ldots, \langle\rho, \varphi_k\rangle\big)\big(-\operatorname{div}(\rho\nabla\varphi_j)\big) \in H_{-1,\rho}\big(\mathbb{R}^d\big). \end{aligned} \tag{1.37}$$

Let

$$f(\rho) := \frac{1}{2} d^2(\rho, \gamma), \quad \text{for some } \gamma \in \mathcal{P}_2(\mathbb{R}^d). \tag{1.38}$$

Let ρ^p be defined according to (1.26), then by Theorem 8.4.7 of [2],

$$\frac{d}{dt} f(\rho(t))\Big|_{t=0} = \int_{\mathbb{R}^d \times \mathbb{R}^d} (x - y)\nabla p(x)\pi_0(dx, dy), \quad \forall \pi_0 \in \Pi^{opt}(\rho, \gamma), \tag{1.39}$$

where $\Pi^{opt}(\rho, \gamma)$ is the collection of optimal transport measures between ρ and γ – see (D.26) on p. 380 of [14] when $c(x, y) = |x - y|^2$. When $\rho(dx) = \rho(x)\,dx$ has Lebesgue density, $\pi_0(dx, dy) = \delta_{\nabla\varphi(x)}(dy)\rho(dx)$ is uniquely determined by the Brenier optimal transport map $T(x) = \nabla\varphi(x)$ in the sense that T pushes measure ρ forward to γ (denoted as $(\nabla\varphi)_\# \rho = \gamma$) where $\varphi = \varphi_{\rho,\gamma}$ is a convex function. Denote

$$p_{\rho,\gamma}(x) = \frac{|x|^2}{2} - \varphi_{\rho,\gamma}(x),$$

then

$$\operatorname{grad} f(\rho) = -\operatorname{div}(\rho\nabla p_{\rho,\gamma}) \in H_{-1,\rho}(\mathbb{R}^d). \tag{1.40}$$

Consequently, for ρ with Lebesgue density,

$$\|\operatorname{grad} f(\rho)\|_{-1,\rho}^2 = d^2(\rho, \gamma). \tag{1.41}$$

See Theorem D.25 on p. 381 of [14].

1.3. Hamilton–Jacobi PDEs in the space of probability measures

In this section, we assume that $\|V \vee 0\|_\infty < \infty$, that $|V| \leqslant \hat{\zeta}(I)$ for some sub-linear continuous function $\hat{\zeta}$ and that V is continuous on finite level sets of I. Condition 1.5 provides a set of requirements on Φ, ϕ and F so that the above properties are always satisfied (see Lemma 2.18).

For $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $n \in H_{-1,\rho}(\mathbb{R}^d)$, we define a Hamiltonian function H through Legendre transform of L,

$$\begin{aligned} H(\rho, n) &:= \sup_{m \in H_{-1,\rho}(\mathbb{R}^d)} \big(\langle n, m\rangle_{-1,\rho} - L(\rho, m)\big) \\ &= \sup_{m \in H_{-1,\rho}(\mathbb{R}^d)} \left(\langle n, m\rangle_{-1,\rho} - \frac{1}{2}\big\|m - \nu\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big)\big\|_{-1,\rho}^2\right) + V(\rho). \end{aligned} \tag{1.42}$$

If ρ satisfies $I(\rho) < \infty$, then $\operatorname{grad} S(\rho) \in H_{-1,\rho}(\mathbb{R}^d)$ and

$$H(\rho, n) = \frac{1}{2}\|n\|_{-1,\rho}^2 + \nu\langle -\operatorname{grad} S(\rho), n\rangle_{-1,\rho} + V(\rho).$$

Natural extension of the above expression to certain class of $n \in \mathcal{D}'(\mathbb{R}^d)$ is straightforward. For instance, in the case of general $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, we can still formally apply $n = \epsilon \operatorname{grad} S(\rho)$ to the above expression

$$H\big(\rho, \epsilon \operatorname{grad} S(\rho)\big) := \left(\frac{\epsilon^2}{2} - \nu\epsilon\right) I(\rho) + V(\rho).$$

Provided $0 < \epsilon < 2\nu$, the right hand side above is upper semicontinuous in $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ in the weak convergence of probability measure topology (i.e. narrow topology), and takes values in $\mathbb{R} \cup \{-\infty\}$. Similarly, for all $\epsilon > 0$,

$$H\big(\rho, -\epsilon \operatorname{grad} S(\rho)\big) := \left(\frac{\epsilon^2}{2} + \nu\epsilon\right) I(\rho) + V(\rho)$$

with a right hand side being lower semicontinuous in $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and taking values in $\mathbb{R} \cup \{+\infty\}$. Moreover, in view of (1.40), for $I(\rho) < \infty$,

$$H\left(\rho, \operatorname*{grad}_{\rho}\left(\frac{\theta}{2}d^2(\rho,\gamma)\right)\right)$$
$$= -\nu\theta \int_{\mathbb{R}^d} \nabla p_{\rho,\gamma} \frac{\nabla\rho + \rho\nabla\Psi}{\rho}\, d\rho + \frac{\theta^2}{2} d^2(\rho,\gamma) + V(\rho).$$

Now, let

$$D_0 := \left\{ f_0(\rho) = \frac{\theta}{2} d^2(\rho,\gamma) + \epsilon S(\rho) + c\colon\ c \in \mathbb{R}, \theta > 0, 0 < \epsilon < 2\nu, \gamma \in \mathcal{P}_2(\mathbb{R}^d) \right\},$$
$$D_1 := \left\{ f_1(\gamma) = -\frac{\theta}{2} d^2(\rho,\gamma) - \epsilon S(\gamma) + c\colon\ c \in \mathbb{R}, \theta > 0, 0 < \epsilon < 2\nu, \rho \in \mathcal{P}_2(\mathbb{R}^d) \right\},$$

and

$$D := D_0 \cup D_1. \tag{1.43}$$

For each $f_0 \in D_0$ and ρ in the effective domain of f_0 (i.e. $S(\rho) < \infty$), by (1.30) and (1.40),

$$\operatorname{grad} f_0(\rho) = -\operatorname{div}\left(\rho\left(\epsilon \frac{\nabla\rho + \rho\nabla\Psi}{\rho} + \theta\nabla p_{\rho,\gamma}\right)\right).$$

Similar relation also holds for $f_1 \in D_1$. With the above discussions, we define operator $H : D \mapsto M(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$ as follows

$$Hf(\rho) := \begin{cases} H(\rho, \operatorname{grad} f(\rho)) & \text{when } I(\rho) < \infty, \\ -\infty & \text{when } I(\rho) = +\infty,\ f \in D_0, \\ +\infty & \text{when } I(\rho) = +\infty,\ f \in D_1. \end{cases} \tag{1.44}$$

In the above, $M(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$ means the collection of all measurable functions from $\mathcal{P}_2(\mathbb{R}^d)$ to $\overline{\mathbb{R}}$. By Lemma 5.1, $Hf_0 : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous for $f_0 \in D_0$ and $Hf_1 : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous for $f_1 \in D_1$.

We are interested in two kinds of Hamilton–Jacobi equations. By resolvent type, we mean

$$f(\rho) - \alpha Hf(\rho) = h(\rho), \quad \rho \in \mathcal{P}_2(\mathbb{R}^d). \tag{1.45}$$

Here $\alpha > 0$ and $h : \mathcal{P}_2(\mathbb{R}^d) \mapsto \overline{\mathbb{R}}$ are given. By Cauchy problem, we mean

$$\begin{cases} \partial_t U(t,\rho) = HU(t,\rho), & (t,\rho) \in (0,T) \times \mathcal{P}_2(\mathbb{R}^d), \\ U(0,\rho) = g(\rho), & \rho \in \mathcal{P}_2(\mathbb{R}^d), \end{cases} \tag{1.46}$$

where the function $g : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}$ is prescribed. The precise meaning of each equation is spelled out in Definitions 1.11 and 1.12 respectively.

In Theorems 1.13 and 1.14 we prove that the above two equations are both well posed and the respective solution is given by

$$U(t,\rho_0) = \sup\left\{ g(\rho(t)) - \int_0^t L(\rho(s), \dot\rho(s))\, ds\colon\ \rho(0) = \rho_0,\ \rho(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d)) \right\}$$
$$= \sup\{ g(\rho_1) - D(\rho_1, \rho_0; t)\colon\ \rho_1 \in \mathcal{P}_2(\mathbb{R}^d) \}, \tag{1.47}$$

and

$$f(\rho_0) = \sup\left\{ \int_0^\infty e^{-\alpha^{-1}s} (\alpha^{-1} h(\rho) - L(\rho, \dot\rho))\, ds\colon\ \rho(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d)), \rho(0) = \rho_0 \right\}. \tag{1.48}$$

The study of Hamilton–Jacobi equation in Banach space was initiated by Crandall and Lions [6,7]. See also there for further references. For the setup in this article, it is more profitable to view the space of probability measures as a metric space. Our proofs rely on new ideas introduced in Feng and Katsoulakis [13], and Feng and Kurtz [14]. Indeed, what were considered in these two references are exactly the special case where $V \equiv 0$. Feng and Katsoulakis [13] only discussed uniqueness. Feng and Kurtz [14] not only discussed the uniqueness result, but also constructed solutions indirectly using a related probabilistic large deviation problem. In particular, it used a type of operator extension technique developed in viscosity solution context. In this article, we consider the general case of $V \not\equiv 0$ given by (1.5) and handle existence and uniqueness together in a direct way using the calculus developed in [2], avoiding the involved viscosity extension method.

1.4. Main results

Throughout this article, we assume that

Condition 1.5.

(1) *$\Phi, \phi \in C^1(\mathbb{R}^d)$, Φ is even $\Phi(-x) = \Phi(x)$, and*

$$\left|\nabla\Phi(x)\right| + \left|\nabla\phi(x)\right| \leqslant A\left(1 + |x|\right)$$

for some constant $A > 0$.

(2) *$\Psi \in C^4(\mathbb{R}^d)$, the Hessian $D^2\Psi(x) \geqslant \lambda_\Psi I$ for some $\lambda_\Psi \in \mathbb{R}$;*

(3) *$\int_{\mathbb{R}^d} \Psi e^{-2\Psi}\,dx < \infty$;*

(4) *Ψ has super-quadratic growth at infinity, i.e.*

$$\lim_{|x|\to\infty} \frac{\Psi(x)}{|x|^2} = +\infty;$$

(5) *there exists an $\omega \in C(\mathbb{R}_+)$ with $\omega(0) = 0$ and at most polynomial growth at infinity such that*

$$\Psi(y) - \Psi(x) \leqslant \omega\left(|y - x|\right)\left(1 + \Psi(x)\right), \tag{1.49}$$

$$\left|\nabla\Psi(y) - \nabla\Psi(x)\right|^2 \leqslant \omega\left(|y - x|\right)\left(1 + \left|\nabla\Psi(x)\right|^2 + \Psi(x)\right); \tag{1.50}$$

(6) *$|\nabla\Psi|^2$ dominates the growth of $\Delta\Psi$ in the sense that: there exists real constants $A, B > 0$ with $A < 1$ such that*

$$2\Delta\Psi \leqslant A|\nabla\Psi|^2 + B; \tag{1.51}$$

(7) *$\psi(x) := |\nabla\Psi|^2 - 2\Delta\Psi$ has super-quadratic growth at infinity:*

$$\lim_{|x|\to\infty} |x|^{-2}\psi(x) = +\infty; \tag{1.52}$$

(8) *$D^2\psi(x) \geqslant \lambda_\psi I$ for some $\lambda_\psi \in \mathbb{R}$;*

(9) *$F \in C^1(\mathbb{R}_+)$, $F(0) = 0$,*

$$\limsup_{r\to 0+} \frac{|F(r)|}{r^\alpha} < +\infty \quad \text{for some} \quad \alpha \in \left(\frac{d}{d+2}, 1\right), \tag{1.53}$$

and

$$\limsup_{r\to+\infty} \frac{|F(r)| + |rF'(r)|}{r^\beta} < +\infty, \tag{1.54}$$

for some $\beta \in [1, \frac{d+2}{d})$ when $d \geqslant 3$, and $\beta \in [1, 2)$ when $d = 1, 2$.

If the leading order terms in $\Psi, \nabla\Psi$ have polynomial growth, then (1.49)–(1.50) are satisfied. Typical examples satisfying all requirements for Ψ (and ψ) are polynomials with leading order term more than quadratic.

Since Ψ has super-quadratic growth, the relative entropy functional S has compact level set property in $\mathcal{P}_2(\mathbb{R}^d)$. Under (1.52), Fisher information functional I in (1.33) has this property as well.

The function $F(r) = Cr \log r$ is a good choice satisfying the conditions for F for all dimensions $d = 1, 2, \ldots$. The lower the dimension d, the broader the class of F is allowed. In particular, in the case of three dimension $d = 3$, $F(r) = Cr^\gamma$ with $\gamma \in [1, \frac{5}{3})$ satisfies the requirements.

Under the assumptions on F, Φ and ϕ in Condition 1.5, the following holds (Lemma 2.18).

Condition 1.6. $V : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{\pm\infty\}$ *satisfies*

(1) *there exists a right continuous, nondecreasing function* $\hat{\zeta} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *with sub-linear growth at infinity*

$$\lim_{r \to \infty} r^{-1} \hat{\zeta}(r) = 0$$

such that

$$\left|V(\rho)\right| \leqslant \hat{\zeta}\big(I(\rho)\big), \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big); \tag{1.55}$$

(2) V *is continuous on finite level sets of* I:

$$\lim_{n \to \infty} V(\rho_n) = V(\rho) \quad \textit{whenever } \rho_n \to \rho \textit{ and } \sup_n I(\rho_n) < \infty.$$

If we strengthen the above requirement, better results can be obtained.

Condition 1.7. $V : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{-\infty\}$ *satisfies*

(1) *there exists a right continuous, nondecreasing function* $\hat{\zeta} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *with sub-linear growth at infinity, such that*

$$\left|V(\rho)\right| \leqslant \hat{\zeta}\big(S(\rho)\big), \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big);$$

(2) V *is continuous on finite level sets of* S:

$$\lim_{n \to \infty} V(\rho_n) = V(\rho) \quad \textit{whenever } \rho_n \to \rho \textit{ and } \sup_n S(\rho_n) < \infty.$$

The above condition is stronger than Condition 1.6 because of a version of mass transport inequality $S(\rho) \leqslant I(\rho) + C_\Psi d^2(\rho, \mu^\Psi)$ (Lemma 5.3), and

$$d^2\big(\rho, \mu^\Psi\big) \leqslant C_1\left(1 + \int_{\mathbb{R}^d} |x|^2 \rho(dx)\right) \leqslant C_2\left(1 + \int_{\mathbb{R}^d} \psi(x)\rho(dx)\right) \leqslant C_2\big(1 + I(\rho)\big).$$

The second inequality above follows by the super-quadratic growth assumption on ψ.

For the resolvent problem (1.45), we also need to impose conditions on h. Noting h can be viewed as part of V by replacing V by $V^h = V + \alpha^{-1} h$ and setting $h = 0$, we only need to consider same type of conditions for V to h. We will refer to them respectively as Conditions 1.6, 1.7 for h with the V replaced by h.

For the Cauchy problem, we assume condition on initial data g as follows:

Condition 1.8. $g \in C(\mathcal{P}_2(\mathbb{R}^d))$, *and*

$$-\hat{\zeta}\big(S(\rho)\big) \leqslant g(\rho) \leqslant \|g \vee 0\|_\infty < \infty, \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big)$$

where $\hat{\zeta} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *is nondecreasing with sub-linear growth at infinity.*

Next, we discuss the meaning of solution to (1.1).

In the case of smooth ρ, as observed by [17],

$$2\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} = \frac{1}{2}|\nabla \log \rho|^2 + \Delta \log \rho,$$

$$2\rho \partial_j \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} = \sum_i \partial_i(\rho \partial_i \partial_j \log \rho) = \partial_j \Delta \rho - \sum_i \partial_i\left(\frac{\partial_i \rho \partial_j \rho}{\rho}\right).$$

Consequently, for $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \varphi\rho\partial_j \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\,dx = -\frac{1}{2}\int_{\mathbb{R}^d} \partial_j\Delta\varphi\,d\rho + \frac{1}{2}\sum_i \int_{\mathbb{R}^d} \partial_i\varphi \frac{\partial_i\rho\partial_j\rho}{\rho}\,dx. \tag{1.56}$$

Such observation motivates us to introduce a notion of weak solution as follows:

Definition 1.9. By a weak solution to system (1.1), we mean

(1) $\rho(\cdot) \in AC(0,T;\mathcal{P}_2(\mathbb{R}^d))$ with

$$S\big(\rho(T)\big) + \int_0^T I\big(\rho(t)\big)\,dt < \infty; \tag{1.57}$$

(2) $u : (0,T)\times\mathbb{R}^d \to \mathbb{R}^d$ is Borel measurable satisfying $u(t) \in L^2_{\nabla,\rho(t)}(\mathbb{R}^d)$ (see (1.13) for definition) a.e. $t \in (0,T)$, and

$$\int_0^T \int_{\mathbb{R}^d} \big|u(t,x)\big|^2 \rho(t,dx)\,dt < \infty, \tag{1.58}$$

and

$$\partial_t\rho + \operatorname{div}(\rho u) = 0$$

in the sense of distribution;

(3) for every $\xi \in C_c^\infty((0,T)\times\mathbb{R}^d;\mathbb{R}^d)$,

$$\begin{aligned}
&\int_0^T\int_{\mathbb{R}^d} u(t,x)\cdot\big(\partial_t\xi(t,x) + (u\cdot\nabla)\xi(t,x)\big)\rho(t,dx)\,dt \\
&\quad + \int_0^T\int_{\mathbb{R}^d} P(\rho)\operatorname{div}\xi\,dx\,dt - \int_0^T\int_{\mathbb{R}^d} \nabla(\phi + \Phi*\rho)\cdot\xi\rho(t,dx)\,dt \\
&\quad + \nu^2\int_0^T\int_{\mathbb{R}^d}\bigg(-\frac{\nabla\rho}{\rho}\cdot D\xi\cdot\frac{\nabla\rho}{\rho} + \Delta\operatorname{div}\xi + \frac{1}{2}\xi\cdot\nabla\psi\bigg)\rho(t,dx)\,dt = 0,
\end{aligned} \tag{1.59}$$

where $D\xi = (\partial_i\xi_j)_{(i,j)}$ is a matrix.

Note that estimates (1.57) and (1.58) ensure the integrability of all the terms in (1.59).

Let P_t be the probability transition semigroup on $\mathcal{P}(\mathbb{R}^d)$ generated by $B = \Delta - \nabla\Psi\cdot\nabla$. Combine results of Lemmas 2.1, 3.2 and Theorem 3.7, we have the following:

Theorem 1.10. *Assume that*

$$D(\rho_1\|\rho_0;T) := \inf_{\pi\in\Gamma(\rho_0,\rho_1)} S(\pi\,\|\,P_{\nu T}\otimes\rho_0) < \infty,$$

and that Condition 1.5 *holds. Then minimizer to* (1.24) *exists. Further assume $S(\rho_0) < \infty$, then any such minimizer (denoting it by $\rho(\cdot)$) is a weak solution to* (1.1) *with* $\sup_{0\leqslant t\leqslant T} S(\rho(t)) < \infty$.

We now discuss Hamilton–Jacobi equations. Let $\bar{\mathbb{R}} = \mathbb{R}\cup\{\pm\infty\}$ denote the space of extended real numbers. If E is a metric space, we denote by $M(E;\bar{\mathbb{R}})$ the space of measurable function on E taking values in $\bar{\mathbb{R}}$.

Definition 1.11 *(Resolvent problem).* Let $f \in M(\mathcal{P}_2(\mathbb{R}^d); \bar{R})$. Suppose that $|f(\rho)| \leqslant \zeta(S(\rho))$ holds for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ for some sub-linear function $\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$ and f is continuous on finite level sets of S.

(1) f is called a viscosity sub-solution to (1.45) if for each $f_0 \in D_0$ (see (1.43)), and for each $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ such that $(f - f_0)(\rho_0) = \sup_{\rho \in \mathcal{P}_2(\mathbb{R}^d)}(f - f_0)(\rho)$,
$$\alpha^{-1}(f - h)(\rho_0) \leqslant H f_0(\rho_0).$$
(2) f is called a super-solution to (1.45) if for each $f_1 \in D_1$, and for each $\rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$ such that $(f_1 - f)(\rho_1) = \sup_{\rho \in \mathcal{P}_2(\mathbb{R}^d)}(f_1 - f)(\rho)$,
$$\alpha^{-1}(f - h)(\rho_1) \geqslant H f_1(\rho_1).$$

If f are both sub- and super-solutions to (1.45), we call it a solution.

Definition 1.12 *(Cauchy problem).* Let $U \in M([0,T] \times \mathcal{P}_2(\mathbb{R}^d); \bar{\mathbb{R}})$. Suppose that $|U(t,\rho)| \leqslant \zeta(S(\rho))$ for all $(t,\rho) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ for some sub-linear function $\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$, and that U is continuous on $[0,T] \times K_L$ where $K_L := \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant L\}$ for each $L < \infty$.

(1) U is called a viscosity sub-solution to (1.46), if for each
$$U_0(t,\rho) = \frac{\alpha}{2}|t - s|^2 + \frac{\theta}{2} d^2(\rho,\gamma) + \epsilon S(\rho) + c, \tag{1.60}$$
where $\gamma \in \mathcal{P}_2(\mathbb{R}^d)$, $\alpha, \theta > 0$, $0 < \epsilon < 2\nu$, $c \in \mathbb{R}$, and for each $(t_0, \rho_0) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ such that
$$(U - U_0)(t_0, \rho_0) = \sup_{(t,\rho) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)} (U - U_0)(t,\rho), \tag{1.61}$$
we have
(a) in the case of $t_0 > 0$,
$$(-\partial_t U_0 + H U_0)(t_0, \rho_0) \geqslant 0; \tag{1.62}$$
(b) in the case of $t_0 = 0$,
$$\limsup_{t \to 0+, \rho' \to \rho_0, S(\rho') \leqslant C} U(t, \rho') \leqslant g(\rho_0),$$
for every $C \in [0, \infty)$.

(2) U is called a super-solution to (1.46), if for each
$$U_1(s,\gamma) = -\frac{\alpha}{2}|t - s|^2 - \frac{\theta}{2} d^2(\rho,\gamma) - \epsilon S(\gamma) + c, \tag{1.63}$$
where $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, $\alpha, \theta > 0$, $0 < \epsilon < 2\nu$, $c \in \mathbb{R}$, and for each $(s_0, \gamma_0) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ such that
$$(U_1 - U)(s_0, \gamma_0) = \sup_{(s,\gamma) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)} (U_1 - U)(s,\gamma), \tag{1.64}$$
we have
(a) in the case of $s_0 > 0$,
$$(-\partial_s U_1 + H U_1)(s_0, \gamma_0) \leqslant 0;$$
(b) in the case of $s_0 = 0$,
$$\liminf_{t \to 0+, \gamma' \to \gamma_0, S(\gamma') \leqslant C} U(t, \gamma') \geqslant g(\gamma_0),$$
for every $C \in [0, \infty)$.

If U are both sub- and super-solutions to (1.46), we call it a solution.

In the above, in view of growth estimate $|f(\rho)| \leqslant \zeta(S(\rho))$, $\epsilon S(\rho) - f(\rho)$ is understood as $+\infty$, when $S(\rho) = +\infty$. We adapt this convention throughout this article. Therefore, $f - f_0$ and $f_1 - f$ are always well defined on $\mathcal{P}_2(\mathbb{R}^d)$. Furthermore, by Lemma B.1, they are both upper semicontinuous. The case of $U - U_0$ and $U_1 - U$ is handled in a similar way.

Theorem 1.13. *Assume that Condition* 1.5 *holds*; *that* $\|V \vee 0\|_\infty < \infty$ *and that* g *satisfies Condition* 1.8.
Then the function U *given by* (1.47) *is*

(1) *a viscosity solution to* (1.46)*, uniquely defined on*

$$[0, \infty) \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d):\ S(\rho) < \infty\},$$

among those satisfying (4.11);
(2) *continuous on* $[0, \infty) \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d):\ S(\rho) \leqslant C\}$ *for every* $C \in \mathbb{R}$;
(3) *the unique continuous viscosity solution on* $[0, T] \times \mathcal{P}_2(\mathbb{R}^d)$*, if Condition* 1.7 *is also satisfied.*

The above results follow from Lemma 4.5, Lemmas 4.7 and 4.8, Lemma 5.6, and Lemmas 5.11 and 5.12.

Theorem 1.14. *Assume that Condition* 1.5 *holds, that* $\|V \vee 0\|_\infty + \|h \vee 0\|_\infty < \infty$*, and that Condition* 1.6 *holds for* h.
Then the value function f *defined by* (1.48) *is*

(1) *a viscosity solution to* (1.45)*, uniquely defined on*

$$\{\rho \in \mathcal{P}_2(\mathbb{R}^d):\ S(\rho) < \infty\},$$

among those satisfying (4.2);
(2) *continuous on every finite level sets of* S;
(3) *the unique continuous viscosity solution on* $\mathcal{P}_2(\mathbb{R}^d)$*, if Condition* 1.7 *is also satisfied for both* h *and* V.

The above results follow from Lemma 4.1, Lemmas 4.3 and 4.4, Lemma 5.5, and Lemmas 5.14 and 5.15.

2. A priori estimates and regularity

2.1. Regularity of paths with finite action

Suppose that $\rho(\cdot) \in C([0, \infty); \mathcal{P}_2(\mathbb{R}^d))$ satisfies the finite action property

$$A_T[\rho(\cdot)] = \int_0^T L(\rho(s), \dot\rho(s))\, ds < \infty.$$

Assuming $\|V \vee 0\|_\infty < \infty$, then

$$\int_0^T T(\rho, \dot\rho)\, ds < \infty. \tag{2.1}$$

Therefore, it follows from Lemma D.34 and Appendix D of Feng and Kurtz [14] that there exists

$$m := -\operatorname{div}(\rho v) \in H_{-1,\rho}(\mathbb{R}^d) \quad \text{with } v \in L^2_{\nabla,\rho}(\mathbb{R}^d)$$

such that the equation

$$\partial_t \rho = \nu\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big) + m \tag{2.2}$$

holds in the Schwartz distributional sense. Consequently, we can write

$$\int_0^T T(\rho,\dot\rho)\,dt = \int_0^T \frac{1}{2}\big\|m(t)\big\|_{-1,\rho(t)}^2\,dt = \int_0^T \frac{1}{2}\int_{\mathbb{R}^d} \big|v(t,x)\big|^2 \rho(t,dx)\,dt < \infty. \tag{2.3}$$

We will prove in this section that

Lemma 2.1. *For the above ρ, we have $S(\rho(r)) < \infty$ for each $r \in (0,T]$. Moreover, $\rho \in AC^2((s,t);\mathcal{P}_2(\mathbb{R}^d))$ for every $0 < s < t \leqslant T$, and*

$$\int_s^T \big(I\big(\rho(r)\big)\,dr + \|\dot\rho\|_{-1,\rho(r)}^2\big)\,dr < \infty.$$

In particular, for $0 < r \leqslant T$,

$$\operatorname{grad} S\big(\rho(r)\big) = -\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big) \in H_{-1,\rho(r)}\big(\mathbb{R}^d\big).$$

Furthermore, for $0 \leqslant s < t \leqslant T$,

$$S\big(\rho(t)\big) + \nu\int_s^t I\big(\rho(r)\big)\,dr \leqslant S\big(\rho(s)\big) + \int_s^t \big\|m(r)\big\|_{-1,\rho(r)}\sqrt{I\big(\rho(r)\big)}\,dr.$$

The main goal of this subsection is to establish the above estimate as a combination of two inequalities. First, we prove an entropy–entropy production estimate (Lemmas 2.5, 2.6). Second, we show a variational inequality (Lemma 2.7). We note that in the case of path generated by gradient flows, these estimates are standard (e.g. Theorems 11.1.4 and 11.1.6 in [2]). The point to emphasize here is that we are handling a much broader class of paths which is not clear to be even absolutely continuous in time *a priori*. Therefore, the calculus of [2] cannot be applied.

There are two ways to look at (2.2). In our variational problem context, the correct way is to view ρ as a given path in $C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ which, by computation, produces m. Therefore, the introduction of m contains no ambiguity. Once m is given, we can also think of (2.2) as a controlled partial differential equation with m as a control. However, extensive care is needed if we want to perturb such controlled equation to produce approximations. Note that m has fairly singular implicit dependence on ρ in (2.2). For instance, the tangent spaces $H_{-1,\rho}(\mathbb{R}^d)$ and $H_{-1,\gamma}(\mathbb{R}^d)$ may have no overlap when $\rho \neq \gamma$. To make sense out of this control interpretation, we embed the above highly state-dependent control into a much larger space where the state dependency disappears. This is essentially the idea of relaxed control in the sense of L.C. Young. We will define a notion of weak solution on one big implicitly defined canonical reference space where (ρ, m) (or equivalently (ρ, v)) becomes just a low-dimensional projection of such a generalized solution. Approximation will be performed in the higher dimensional generalized solution space. The tool we use to make this clear is probability theory.

If we can assume that Ψ is at most quadratic growth at infinity, then all the delicate estimates using probability can be avoided by direct mollification based deterministic analysis method. But in this article, we needed Ψ (as well as ψ) to be super-quadratic at infinity, so that, among other things, relative entropy S and relative Fisher information I have compact level sets in $\mathcal{P}_2(\mathbb{R}^d)$ with Wasserstein 2-metric.

By an argument of Kurtz and Stockbridge [19] (see Section 13.3.5, especially p. 340 of Feng and Kurtz [14] for an explanation), there exists a pair of stochastic processes $\{(X(t),\Lambda(t))\colon t \geqslant 0\}$ with values in $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ such that, for every $r \geqslant 0$ and every $\varphi \in C_b(\mathbb{R}^d \times \mathbb{R}^d)$,

$$E\bigg[\int_{\mathbb{R}^d} \varphi\big(X(r),v\big)\Lambda(r,dv)\bigg] = E\big[\varphi(X(r),v\big(r,X(r)\big))\big] = \int_{\mathbb{R}^d} \varphi\big(x,v(r,x)\big)\rho(r,dx);$$

moreover,

$$X(t) = X(0) + \int_0^t \big(-\nu\nabla\Psi\big(X(r)\big) + U(r)\big)\,dr + \sqrt{2\nu}\,W(t), \tag{2.4}$$

where W is a d-dimensional standard Brownian motion and $U(r)=\int_{\mathbb{R}^d} z\Lambda(r,dz)$. It follows then

$$E\big[\big|U(r)\big|^2\big]\leqslant E\bigg[\int\limits_{\mathbb{R}^d}|v|^2\Lambda(r,dv)\bigg]=\int\limits_{\mathbb{R}^d}\big|v(r,x)\big|^2\rho(r,dx)=\big\|m(r)\big\|^2_{-1,\rho(r)}.$$

Weak solution to stochastic differential equation (2.4) refers to a probability law

$$P\big((X,\Lambda)\in\cdot\big)\in\mathcal{P}\big(C\big([0,\infty);\mathbb{R}^d\times\mathcal{P}\big(\mathbb{R}^d\big)\big)\big).$$

Trajectory

$$\rho(t,dx)=P\big(X(t)\in dx\big)$$

is a reduced level description, in the sense that it is just a projection of the P at the X component at time t.

Let $0\leqslant s\leqslant t\leqslant T$ be given. We consider stochastic representation (2.4) of (2.2) but with initial value start at time $s\geqslant 0$

$$X(t)=X(s)+\int\limits_s^t\big(-\nu\nabla\Psi\big(X(r)\big)+U(r)\big)\,dr+\sqrt{2\nu}\big(W(t)-W(s)\big). \tag{2.5}$$

Set stopping time

$$\tau_k=\inf\big\{r\geqslant s\colon\ \big|X(r)\big|\geqslant k\big\}.$$

Note that $\rho(s)\in\mathcal{P}_2(\mathbb{R}^d)$ (i.e. $E[|X(s)|^2]<\infty$). Pages 340–341 of Feng and Kurtz [14] give the following estimate for some super-linear function $\eta:\mathbb{R}_+\mapsto\mathbb{R}_+$:

$$\sup_k\sup_{s\leqslant r\leqslant T}E\big[\eta\big(\big|X(r\wedge\tau_k)\big|^2\big)\big]\leqslant C_T<\infty,\quad\forall T>0.$$

Therefore, for $t\geqslant s$,

$$P(\tau_k\leqslant t)\leqslant P\big(\big|X(t\wedge\tau_k)\big|\geqslant k\big)\leqslant k^{-2}E\big[\big|X(t\wedge\tau_k)\big|^2\big]\leqslant k^{-2}C,\quad\text{for some } 0<C<\infty;$$

$\lim_{k\to\infty}\tau_k=+\infty$ in probability. Since $s\leqslant\cdots\leqslant\tau_k\leqslant\tau_{k+1}\leqslant\cdots$, by monotone convergence, $\lim_{k\to\infty}\tau_k=+\infty$ almost surely. Assume $E[\Psi(X(s))]<\infty$, by Ito's formula,

$$E\big[\Psi\big(X(t\wedge\tau_k)\big)\big]=E\big[\Psi\big(X(s)\big)\big]+E\Bigg[\int\limits_s^{t\wedge\tau_k}\big(-\nu\big|\nabla\Psi\big(X(r)\big)\big|^2+U(r)\nabla\Psi\big(X(r)\big)+\nu\Delta\Psi\big(X(r)\big)\big)\,dr\Bigg].$$

Note that

$$|\nabla\Psi|^2-\Delta\Psi=\frac{1}{2}\psi+\frac{1}{2}|\nabla\Psi|^2.$$

Taking $k\to\infty$, by Fatou's lemma and Young's inequality,

$$E\Bigg[\Psi\big(X(t)\big)+\frac{\nu}{2}\int\limits_s^t\bigg(\psi\big(X(r)\big)+\frac{1}{2}\big|\nabla\Psi\big(X(r)\big)\big|^2\bigg)\,dr\Bigg]\leqslant E\Bigg[\Psi\big(X(s)\big)+C_\nu\int\limits_s^t\big|U(r)\big|^2\,dr\Bigg]$$

which implies the following:

$$\int\limits_{\mathbb{R}^d}\Psi\,d\rho(t)+\frac{\nu}{2}\int\limits_s^t\int\limits_{\mathbb{R}^d}\bigg(\psi+\frac{1}{2}|\nabla\Psi|^2\bigg)\,d\rho(r)\,dr\leqslant\int\limits_{\mathbb{R}^d}\Psi\,d\rho(s)+C_\nu\int\limits_s^t\big\|m(r)\big\|^2_{-1,\rho(r)}\,dr. \tag{2.6}$$

In the above, we allow the right hand side to be $+\infty$.

We introduce another approximation to X. We want the approximating processes to have smooth densities, so that the usual deterministic as well as stochastic calculus apply when densities are used as part of test functions.

For each of the estimates below, we will be concerned with only the portion of $\rho(r)$ when $0 \leqslant s \leqslant r \leqslant t$. For notational simplicity, we will treat s as the initial value of time and modify the definition of X by setting $X(r) = X(s)$ when to $r < s$. This implies in particular $\rho(r) = \rho(s)$, $r < s$. Let J denote the standard one-dimensional mollifier with a compact support:

$$J(r) := C \exp\left\{\frac{1}{r^2 - 1}\right\}, \quad \text{for } |r| < 1, \qquad J(r) := 0, \quad \text{for } |r| \geqslant 1,$$

where C is such that J becomes a probability density. $J \in C_c^\infty(\mathbb{R})$ is even. With a slight abuse of notation, we denote

$$J(x) := J(|x|) \in C_c^\infty(\mathbb{R}^d), \qquad J_\delta(z) := \delta^{-d} J(\delta^{-1} z).$$

Let Z be a $\mathbb{R}^d$-valued random variable with probability density $J(z)$, $|Z| \leqslant 1$, and we choose Z such that $\{X(\cdot), Z\}$ are independent. We now define

$$X_\delta(t) := X(t) + \sqrt{\delta} Z.$$

It follows that the following convergence holds almost surely when the continuous trajectory processes are restricted to $[0, \infty)$ (that is, they are viewed as $C([0, \infty); \mathbb{R}^d)$-valued random variables)

$$\lim_{\delta \to 0+} X_\delta = X.$$

We will make use of the following properties regarding one-dimensional (in time) probability densities. Let

$$\rho_\delta(r, dx) := P(X_\delta(r) \in dx).$$

Then by a basic probability result, $\rho_\delta(r, x) = (J_\delta *_x \rho(r, \cdot))(x)$. It follows

$$\rho_\delta(r, \cdot) \in C^\infty(\mathbb{R}^d), \quad r \geqslant s.$$

Lemma 2.2. *Allowing the possibility of $+\infty = +\infty$, we have for each $s \geqslant 0$,*

$$\lim_{\delta \to 0+} \int_{\mathbb{R}^d} \Psi \, d\rho_\delta(s) = \lim_{\delta \to 0+} E[\Psi(X_\delta(s))] = E[\Psi(X(s))] = \int_{\mathbb{R}^d} \Psi \, d\rho(s).$$

Proof. By Fatou's lemma,

$$\liminf_{\delta \to 0+} E[\Psi(X_\delta(s))] \geqslant E[\Psi(X(s))].$$

In particular, if $\int_{\mathbb{R}^d} \Psi \, d\rho(s) = +\infty$, the conclusion follows.

Therefore, we only need to prove

$$\limsup_{\delta \to 0+} E[\Psi(X_\delta(s))] \leqslant E[\Psi(X(s))]$$

under the assumption of $\int_{\mathbb{R}^d} \Psi \, d\rho(s) < \infty$. The above follows by the assumed (1.49) on Ψ, and by independence of Z and X: when δ is small enough,

$$E[\Psi(X_\delta(s))] \leqslant E[\Psi(X(s))] + E[\omega(\sqrt{\delta} Z)](1 + E[\Psi(X(s))]). \qquad \square$$

Lemma 2.3.

$$\lim_{\delta \to 0+} S(\rho_\delta(s)) = S(\rho(s)).$$

Proof. By variational representation (1.27), S is lower semicontinuous with respect to weak convergence of probability measure (i.e. the narrow) topology. Therefore we only need to prove

$$\limsup_{\delta \to 0+} S(\rho_\delta(s)) \leqslant S(\rho(s)).$$

We only need to deal with the case of $S(\rho(s)) < \infty$, which we assume from this point on.

Let $\delta > 0$ and recall that $\rho_\delta = J_\delta *_x \rho$. We note that

$$\int_{\mathbb{R}^d} \rho \log \rho \, dx + \int_{\mathbb{R}^d} \Psi \, d\rho = \sup_{f \in C_b(\mathbb{R}^d)} \left\{ \langle f, \rho \rangle - \log \int_{\mathbb{R}^d} e^{f-\Psi} \, dx \right\}$$
$$= \sup_{g = f - \Psi, f \in C_b(\mathbb{R}^d)} \left\{ \langle g, \rho \rangle - \log \int_{\mathbb{R}^d} e^g \, dx \right\} + \int_{\mathbb{R}^d} \Psi \, d\rho.$$

By further approximation, for $\int_{\mathbb{R}^d} \Psi \, d\rho < \infty$,

$$\int_{\mathbb{R}^d} \rho \log \rho \, dx = \sup \left\{ \langle g, \rho \rangle - \log \int_{\mathbb{R}^d} e^g \, dx\colon \; g \in C\big(\mathbb{R}^d\big), \sup_x g < \infty, \int_{\mathbb{R}^d} e^g \, dx < \infty \right\}.$$

On the other hand, for any function g defined as above, by Jensen's inequality and translation invariance property of the Lebesgue measure,

$$\log \int_{\mathbb{R}^d} e^{J_\delta * g} \, dx \leqslant \log \int_{\mathbb{R}^d} e^g \, dx.$$

This implies that

$$\big\langle g, J_\delta * \rho(s) \big\rangle - \log \int_{\mathbb{R}^d} e^g \, dx \leqslant \big\langle J_\delta * g, \rho(s) \big\rangle - \log \int_{\mathbb{R}^d} e^{J_\delta * g} \, dx \leqslant \int_{\mathbb{R}^d} \rho(s,x) \log \rho(s,x) \, dx.$$

Hence

$$\limsup_{\delta \to 0+} \int_{\mathbb{R}^d} \rho_\delta(s,x) \log \rho_\delta(s,x) \, dx \leqslant \int_{\mathbb{R}^d} \rho(s,x) \log \rho(s,x) \, dx.$$

Combined with the result in Lemma 2.2, we have

$$\limsup_{\delta \to 0+} S\big(\rho_\delta(s)\big) \leqslant S\big(\rho(s)\big). \qquad \square$$

Lemma 2.4. *Let $\rho \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ satisfy (2.2) with (2.3). Let $0 \leqslant s \leqslant t \leqslant T$ and suppose further that $\int_{\mathbb{R}^d} \Psi \, d\rho(s) < \infty$. Then*

$$\|\partial_r \rho_\delta\|_{L^\infty((s,t)\times\mathbb{R}^d)} + \|\partial_r \rho_\delta\|_{L^1((s,t)\times\mathbb{R}^d)} < \infty,$$
$$\|\Delta \rho_\delta\|_{L^\infty((s,t)\times\mathbb{R}^d)} + \|\Delta \rho_\delta\|_{L^1((s,t)\times\mathbb{R}^d)} < \infty.$$

Proof. Note that

$$\partial_t \rho_\delta = \nu \Delta \rho_\delta + \nu \nabla\big(J_\delta * (\rho \nabla \Psi)\big) + J_\delta * m.$$

First, for $|r| < \delta$,

$$\frac{1}{|(r/\delta)^2 - 1|^k} e^{\frac{1}{(r/\delta)^2 - 1}} \leqslant C_{1,k,\delta} e^{\frac{1}{2}\frac{1}{(r/\delta)^2-1}} \leqslant C_{2,k,\delta} e^{\frac{1}{(r/2\delta)^2-1}}, \quad k = 1, 2, \ldots,$$

where the last inequality above follows because of

$$\frac{1}{2} \frac{1}{(r/\delta)^2 - 1} \leqslant \frac{1}{(r/2\delta)^2 - 1} + \frac{1}{2}.$$

Therefore, there exists finite constant $C_\delta > 0$ such that

$$|\nabla J_\delta| + |\Delta J_\delta| \leqslant C_\delta J_{2\delta}.$$

By Young inequality for convolution, for both $p = 1$ and $+\infty$,

$$\|\Delta\rho_\delta\|_{L^p((s,t)\times\mathbb{R}^d)} \leqslant \|\Delta J_\delta\|_{L^p((s,t)\times\mathbb{R}^d)}\|\rho\|_{L^1((s,t)\times\mathbb{R}^d)} \leqslant C_\delta\|J_{2\delta}\|_{L^p((s,t)\times\mathbb{R}^d)}|t-s| < \infty,$$

$$\begin{aligned}\big\|J_\delta * \nabla(\rho\nabla\Psi)\big\|_{L^p((s,t)\times\mathbb{R}^d)} &\leqslant \|\nabla J_\delta\|_{L^p((s,t)\times\mathbb{R}^d)}\|\rho\nabla\Psi\|_{L^1((s,t)\times\mathbb{R}^d)} \\ &\leqslant C_\delta\|J_{2\delta}\|_{L^p((s,t)\times\mathbb{R}^d)}\int_s^t\bigg(\int_{\mathbb{R}^d}|\nabla\Psi|^2\,d\rho(r,x)\,dx\bigg)^{1/2}dr < \infty,\end{aligned}$$

$$\begin{aligned}\big\|J_\delta * \nabla(\rho v)\big\|_{L^p((s,t)\times\mathbb{R}^d)} &\leqslant \|\nabla J_\delta\|_{L^p((s,t)\times\mathbb{R}^d)}\|\rho v\|_{L^1((s,t)\times\mathbb{R}^d)} \\ &\leqslant C_\delta\|J_{2\delta}\|_{L^p((s,t)\times\mathbb{R}^d)}\int_s^t\bigg(\int_{\mathbb{R}^d}\big|v(r,x)\big|^2\,d\rho(r,x)\,dx\bigg)^{1/2}dr < \infty,\end{aligned}$$

where we used the fact that $m = -\nabla(\rho v)$ and the estimate (2.6) to get

$$\int_s^t\int_{\mathbb{R}^d}|\nabla\Psi|^2\,d\rho\,dr < \infty. \qquad \square$$

Lemma 2.5. *For $0 \leqslant s < t \leqslant T$, we have*

$$S\big(\rho(t)\big) + \frac{\nu}{2}\int_s^t I\big(\rho(r)\big)\,dr \leqslant S\big(\rho(s)\big) + C_\nu\int_0^t\big\|m(r)\big\|^2_{-1,\rho(r)}\,dr.$$

Proof. We only need to prove the case when the right hand side of the inequality is finite.

We assumed that $0 \leqslant S(\rho(s)) < \infty$. Since $\Psi(x)$ exhibits super-quadratic growth as $|x| \to \infty$, and since $\rho(s) \in \mathcal{P}_2(\mathbb{R}^d)$ has finite second moment, $\rho(s,dx) = \rho(s,x)\,dx$ has Lebesgue density and $|\int_{\mathbb{R}^d}\rho(s,x)\log\rho(s,x)\,dx| < \infty$. This in turn also implies that

$$E\big[\Psi\big(X(s)\big)\big] = \int_{\mathbb{R}^d}\Psi(x)\rho(s,dx) < \infty.$$

Therefore we immediately have the estimate in (2.6).

Observe that

$$X_\delta(t) = X_\delta(s) + \int_s^t\big(-\nu\nabla\Psi\big(X(r)\big) + U(r)\big)\,dr + \sqrt{2\nu}\int_s^t dW(r),$$

and $X(r) = X_\delta(r) - \sqrt{\delta}Z$. Define

$$f(t,x) := \log\big((\rho_\delta + \epsilon)e^{\Psi}\big)(t,x),$$

and denote

$$\nabla_\Psi\varphi := e^{\Psi}\nabla\big(e^{-\Psi}\varphi\big), \qquad \Delta_\Psi\varphi := \operatorname{div}_\Psi\nabla\varphi = e^{\Psi}\nabla\big(e^{-\Psi}\nabla\varphi\big) = (\Delta - \nabla\Psi\nabla)\varphi.$$

Then $f(t,\cdot) \in C^4(\mathbb{R}^d)$, $\epsilon > 0$. To simplify, we write $\rho = \rho_\delta$,

$$\nabla f = \frac{\nabla\rho}{\rho+\epsilon} + \nabla\Psi, \qquad \partial_r f = \frac{\partial_r\rho}{\rho+\epsilon},$$

and

$$\begin{aligned}
\Delta_\Psi f &= \frac{\Delta\rho}{\rho+\epsilon} - \left|\frac{\nabla\rho}{\rho+\epsilon}\right|^2 - |\nabla\Psi|^2 - \nabla\Psi\frac{\nabla\rho}{\rho+\epsilon} + \Delta\Psi \\
&= \frac{\Delta\rho}{\rho+\epsilon} - \frac{1}{2}\left|\frac{\nabla\rho}{\rho+\epsilon} + \nabla\Psi\right|^2 - \frac{1}{2}\left|\frac{\nabla\rho}{\rho+\epsilon}\right|^2 - \frac{1}{2}|\nabla\Psi|^2 + \Delta\Psi \\
&= \frac{\Delta\rho}{\rho+\epsilon} - \frac{1}{2}\left|\frac{\nabla\rho}{\rho+\epsilon} + \nabla\Psi\right|^2 - \frac{1}{2}\left|\frac{\nabla\rho}{\rho+\epsilon}\right|^2 - \frac{1}{2}\psi.
\end{aligned}$$

Let

$$\tau_{k,\delta} = \inf\bigl\{t \geqslant s\colon \bigl|X_\delta(t)\bigr| > k\bigr\}.$$

We have $\lim_{k\to\infty}\tau_{k,\delta} = +\infty$ almost surely, for each $\delta > 0$ fixed. By Ito's formula and Young's inequality,

$$\begin{aligned}
E\bigl[f\bigl(t, X_\delta(t\wedge\tau_{k,\delta})\bigr)\bigr] &= E\left[f\bigl(s, X_\delta(s)\bigr) + \int_s^{t\wedge\tau_{k,\delta}} \bigl(\partial_r f + \nu\Delta_\Psi f + U(r)\nabla f\bigr)\bigl(r, X_\delta(r)\bigr)\,dr\right] \\
&\quad + \nu E\left[\int_s^{t\wedge\tau_{k,\delta}} \bigl(\nabla\Psi\bigl(X_\delta(r)\bigr) - \nabla\Psi\bigl(X(r)\bigr)\bigr)\nabla f\bigl(r, X_\delta(r)\bigr)\,dr\right] \\
&\leqslant E\bigl[f\bigl(s, X_\delta(s)\bigr)\bigr] - \frac{\nu}{2}E\left[\int_s^{t\wedge\tau_{k,\delta}} \left(\left|\frac{\nabla\rho}{\rho+\epsilon}\right|^2 + \psi\right)\bigl(r, X_\delta(r)\bigr)\,dr\right] \\
&\quad + C_\nu E\left[\int_s^{t\wedge\tau_{k,\delta}} \bigl|U(r)\bigr|^2 + \bigl|\nabla\Psi\bigl(X_\delta(r)\bigr) - \nabla\Psi\bigl(X(r)\bigr)\bigr|^2\,dr\right] \\
&\quad + E\left[\int_s^{t\wedge\tau_{k,\delta}} \frac{\partial_r\rho}{\rho+\epsilon}\bigl(r, X_\delta(r)\bigr)\,dr\right] + \nu E\left[\int_s^{t\wedge\tau_{k,\delta}} \frac{\Delta\rho}{\rho+\epsilon}\bigl(r, X_\delta(r)\bigr)\,dr\right].
\end{aligned} \tag{2.7}$$

We claim that the last two terms are zero in the limit $\lim_{\epsilon\to0+}\lim_{k\to\infty}$. We prove this next. We write ρ_δ instead of ρ to emphasize the dependence on δ again. First, by Lemma 2.4,

$$\sup_k E\left[\int_s^{t\wedge\tau_{k,\delta}} \left|\frac{\partial_r\rho_\delta}{\rho_\delta+\epsilon}\bigl(r, X_\delta(r)\bigr)\right|^2 dr\right] \leqslant \epsilon^{-2}\|\partial_r\rho_\delta\|^2_{L^\infty((s,t)\times\mathbb{R}^d)}(t-s) < \infty.$$

Consequently, by uniform integrability,

$$\lim_{k\to\infty} E\left[\int_s^t 1_{r\leqslant t\wedge\tau_{k,\delta}} \frac{\partial_r\rho_\delta}{\rho_\delta+\epsilon}\bigl(r, X_\delta(r)\bigr)\,dr\right] = E\left[\int_s^t \frac{\partial_r\rho_\delta}{\rho_\delta+\epsilon}\bigl(r, X_\delta(r)\bigr)\,dr\right].$$

Note that

$$E\left[\int_s^t \frac{\partial_r\rho_\delta}{\rho_\delta+\epsilon}\bigl(r, X_\delta(r)\bigr)\,dr\right] = \int_s^t\int_{\mathbb{R}^d} \partial_r\rho_\delta\frac{\rho_\delta}{\rho_\delta+\epsilon}\,dx\,dr.$$

Since $\|\partial_r\rho_\delta\|_{L^1((s,t)\times\mathbb{R}^d)} < \infty$ (Lemma 2.4) and $0 \leqslant \rho_\delta/(\rho_\delta+\epsilon) \leqslant 1$, by dominated convergence theorem,

$$\lim_{\epsilon\to0+}\int_s^t\int_{\mathbb{R}^d} \frac{\rho_\delta}{\rho_\delta+\epsilon}(r,x)\partial_r\rho_\delta(r,x)\,dx\,dr = \int_s^t\int_{\mathbb{R}^d} \partial_r\rho_\delta(r,x)\,dx\,dr.$$

Therefore

$$\lim_{\epsilon\to 0+} E\left[\int_s^t \frac{\partial_r \rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = \lim_{\epsilon\to 0+}\int_s^t\int_{\mathbb{R}^d} \frac{\partial_r \rho_\delta}{\rho_\delta+\epsilon}(r,x)\rho_\delta\,dx\,dr = \int_{\mathbb{R}^d}\int_s^t (\partial_r\rho_\delta)\,dr\,dx$$
$$= \int_{\mathbb{R}^d} \rho_\delta(t,x)\,dx - \int_{\mathbb{R}^d}\rho_\delta(s,x)\,dx = 0.$$

In summary,

$$\lim_{\epsilon\to 0+}\lim_{k\to\infty} E\left[\int_s^{t\wedge\tau_{k,\delta}} \frac{\partial_r \rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = 0.$$

Similarly, by Lemma 2.4,

$$\sup_k E\left[\int_s^{t\wedge\tau_{k,\delta}} \left|\frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\right|^2 dr\right] \leqslant \epsilon^{-2}\|\Delta\rho_\delta\|^2_{L^\infty((s,t)\times\mathbb{R}^d)}|t-s| < \infty,$$

implying that

$$\lim_{k\to\infty} E\left[\int_s^{t\wedge\tau_{k,\delta}} \frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = E\left[\int_s^t \frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right].$$

Note also that

$$E\left[\int_s^t \frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = \int_s^t\int_{\mathbb{R}^d} \frac{\rho_\delta}{\rho_\delta+\epsilon}(r,y)\Delta\rho_\delta(r,y)\,dy\,dr,$$

and that $\|\Delta\rho_\delta\|_{L^1((s,t)\times\mathbb{R}^d)} < \infty$, by Lemma 2.4,

$$\lim_{\epsilon\to 0+} E\left[\int_s^t \frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = \lim_{\epsilon\to 0+}\int_s^t\int_{\mathbb{R}^d} \frac{\rho_\delta}{\rho_\delta+\epsilon}(r,y)\Delta\rho_\delta(r,y)\,dy\,dr$$
$$= \int_s^t\int_{\mathbb{R}^d} \Delta\rho_\delta(r,y)\,dy\,dr = 0.$$

The last step above follows by integration by parts and the fact that (recall $|\nabla J_\delta| \leqslant C_\delta J_{2\delta}$)

$$\big|\nabla\rho_\delta(r,x)\big| \leqslant C_\delta(J_{2\delta} *_x \rho)(r,x) \to 0 \quad \text{as } |x|\to\infty.$$

In summary, for each $\delta > 0$ fixed,

$$\lim_{\epsilon\to 0+}\lim_{k\to\infty} E\left[\int_s^{t\wedge\tau_{k,\delta}} \frac{\Delta\rho_\delta}{\rho_\delta+\epsilon}\big(r, X_\delta(r)\big)\,dr\right] = 0.$$

Note that

$$E\big[f\big(t, X_\delta(t\wedge\tau_{k,\delta})\big)\big] = E\big[\log\big(\rho_\delta\big(t, X_\delta(t\wedge\tau_{k,\delta})\big)+\epsilon\big)\big] + E\big[\Psi\big(X_\delta(t\wedge\tau_{k,\delta})\big)\big].$$

By Fatou's lemma and monotone convergence theorem

$$\liminf_{\epsilon\to 0^+}\liminf_{k\to\infty} E\big[f\big(t, X_\delta(t\wedge\tau_{k,\delta})\big)\big] \geqslant E\big[\log\big(\rho_\delta\big(t, X_\delta(t)\big)\big) + \Psi\big(X_\delta(t)\big)\big] = S\big(\rho_\delta(t)\big) - \log Z.$$

Therefore (2.7) simplifies to

$$\begin{aligned}S\big(\rho_\delta(t)\big) \leqslant S\big(\rho_\delta(s)\big) - \frac{\nu}{2}E\Bigg[\int_s^t \bigg(\bigg|\frac{\nabla\rho_\delta}{\rho_\delta}\bigg|^2 + \psi\bigg)\big(r, X_\delta(r)\big)\,dr\Bigg] \\ + C_\nu \int_s^t \big\|m(r)\big\|^2_{-1,\rho(r)}\,dr \\ + C_\nu E\Bigg[\int_s^t \big|\nabla\Psi\big(X_\delta(r)\big) - \nabla\Psi\big(X(r)\big)\big|^2\,dr\Bigg].\end{aligned}$$

By (1.50) on Ψ, by independence of X and Z, and by the estimates in (2.6),

$$\begin{aligned}&\lim_{\delta\to 0+} E\Bigg[\int_s^t \big|\nabla\Psi\big(X_\delta(r)\big) - \nabla\Psi\big(X(r)\big)\big|^2\,dr\Bigg] \\ &\quad\leqslant \lim_{\delta\to 0+} E\big[\omega(\sqrt{\delta}Z)\big]E\Bigg[\int_s^t \big(1 + \big|\nabla\Psi\big(X(r)\big)\big|^2 + \Psi\big(X(r)\big)\big)\,dr\Bigg] \\ &\quad\leqslant \lim_{\delta\to 0+} E\big[\omega(\sqrt{\delta}Z)\big]\int_s^t \int_{\mathbb{R}^d} \big(1 + |\nabla\Psi|^2 + \Psi\big)\rho(r,dx)\,dr = 0.\end{aligned}$$

Taking $\limsup_{\delta\to 0+}$, by lower semicontinuity of S and I in weak convergence of probability measure topology (i.e. the narrow topology) in ρ,

$$S\big(\rho(t)\big) \leqslant \limsup_{\delta\to 0} S\big(\rho_\delta(s)\big) - \frac{\nu}{2}E\Bigg[\int_s^t \bigg(\bigg|\frac{\nabla\rho}{\rho}\bigg|^2 + \psi\bigg)\,dr\Bigg] + C_\nu \int_s^t \big\|m(r)\big\|^2_{-1,\rho(r)}\,dr.$$

Noting the result of Lemma 2.3, we conclude the lemma. □

Lemma 2.6. *Let $s \geqslant 0$. Assume $S(\rho(s)) < \infty$ and* (2.1) *holds. Then*

$$\int_s^T I\big(\rho(r)\big)\,dr < \infty \tag{2.8}$$

and

$$\int_s^T \|\dot\rho\|^2_{-1,\rho(r)}\,dr < \infty. \tag{2.9}$$

Consequently, $\rho \in AC^2(s,T;\mathcal{P}_2(\mathbb{R}^d))$, and

$$\begin{aligned}&\frac{1}{2}\int_s^T \big\|\dot\rho + \nu\,\mathrm{grad}\,S(\rho)\big\|^2_{-1,\rho}\,dr \\ &\quad= \frac{1}{2}\int_s^T \|\dot\rho\|^2_{-1,\rho}\,dr + \frac{\nu^2}{2}\int_s^T I\big(\rho(r)\big)\,dr + \nu\big(S\big(\rho(T)\big) - S\big(\rho(s)\big)\big).\end{aligned} \tag{2.10}$$

Proof. Lemma 2.5 gives (2.8). Recall $I(\rho) = \|\operatorname{grad} S(\rho)\|_{-1,\rho}^2$, since

$$\|\dot\rho\|_{-1,\rho} \leqslant \big\|\dot\rho + \nu \operatorname{grad} S(\rho)\big\|_{-1,\rho} + \nu\big\|\operatorname{grad} S(\rho)\big\|_{-1,\rho},$$

(2.9) follows and $\rho \in AC^2((s,T);\mathcal{P}(\mathbb{R}^d))$.

Given the estimates above, we have almost everywhere in $t > 0$,

$$\begin{aligned}\big\|\dot\rho + \nu \operatorname{grad} S(\rho)\big\|_{-1,\rho}^2 &= \|\dot\rho\|_{-1,\rho}^2 + \nu^2\big\|\operatorname{grad} S(\rho)\big\|_{-1,\rho}^2 + 2\nu\big\langle\dot\rho, \operatorname{grad} S(\rho)\big\rangle_{-1,\rho}\\ &= \|\dot\rho\|_{-1,\rho}^2 + \nu^2 I(\rho) + 2\nu\frac{d}{dt}S(\rho),\end{aligned}$$

where we applied the chain rule of Proposition 10.3.18 of [2]. (2.10) is an integral version of the above identity. □

Lemma 2.7. *Let $\rho \in C([0,T];\mathcal{P}_2(\mathbb{R}^d))$ satisfy (2.1). For all $\gamma \in \mathcal{P}_2(\mathbb{R}^d)$ with $S(\gamma) < \infty$, and $0 \leqslant s \leqslant t$,*

$$\begin{aligned}&\frac{1}{2}d^2\big(\rho(t),\gamma\big) + \nu\int_s^t S\big(\rho(r)\big)\,dr\\ &\quad\leqslant \frac{1}{2}d^2\big(\rho(s),\gamma\big) + \nu S(\gamma)(t-s) + \int_s^t \left(-\nu\frac{\lambda_\Psi}{2}d^2\big(\rho(r),\gamma\big) + d\big(\rho(r),\gamma\big)\big\|m(r)\big\|_{-1,\rho(r)}\right)dr. \qquad (2.11)\end{aligned}$$

Proof. The proof is divided into two parts.

In the first part, we assume that $\int_{\mathbb{R}^d}\Psi\,d\rho(s) < +\infty$.

From (2.1), we have (2.2). Let $G(z)$ be the probability density for standard normal random variable and let $G_\delta(z) = \delta^{-d}G(\delta^{-1}z)$. Take $\rho_\delta(r) = G_\delta *_x \rho(r)$, and

$$u_\delta(r,x) := \frac{G_\delta *_x (\rho(r)\nabla\Psi)}{\rho_\delta(r)}(x), \qquad v_\delta(r,x) := \frac{G_\delta *_x (\rho(r)v(r))}{\rho_\delta(r)}(x).$$

We have

$$\partial_t\rho_\delta = \nu\big(\Delta\rho_\delta + \operatorname{div}(\rho_\delta u_\delta)\big) - \operatorname{div}(\rho_\delta v_\delta).$$

Therefore

$$\begin{aligned}&\big\|\partial_r\rho_\delta - \nu\big(\Delta\rho_\delta + \operatorname{div}(\rho_\delta\nabla\Psi)\big)\big\|_{-1,\rho_\delta}\\ &\quad\leqslant \big\|\nu\operatorname{div}(\rho_\delta u_\delta)\big\|_{-1,\rho_\delta} + \big\|\operatorname{div}(\rho_\delta v_\delta)\big\|_{-1,\rho_\delta} + \big\|\nu\operatorname{div}(\rho_\delta\nabla\Psi)\big\|_{-1,\rho_\delta}\\ &\quad= \nu\|u_\delta\|_{L^2_{\rho_\delta}} + \|v_\delta\|_{L^2_{\rho_\delta}} + \nu\|\nabla\Psi\|_{L^2_{\rho_\delta}} \leqslant \nu\|\nabla\Psi\|_{L^2_\rho} + \|v\|_{L^2_\rho} + \nu\|\nabla\Psi\|_{L^2_{\rho_\delta}}\end{aligned}$$

where the last line follows from Lemma 8.1.9 of [2]. By (1.50) and the estimates in (2.6),

$$\lim_{\delta\to 0+}\int_s^T\int_{\mathbb{R}^d}|\nabla\Psi|^2\,d\rho_\delta(r)\,dr = \int_s^T\int_{\mathbb{R}^d}|\nabla\Psi|^2\,d\rho(r)\,dr < \infty.$$

Consequently

$$\int_s^T\big\|\partial_r\rho_\delta - \nu\big(\Delta\rho_\delta + \operatorname{div}(\rho_\delta\nabla\Psi)\big)\big\|_{-1,\rho_\delta}^2\,dr < \infty.$$

On the other hand, by Jensen's inequality, for any $r \geqslant s$,

$$-c_{\delta,r}\big(1+|x|^2\big) \leqslant \int_{\mathbb{R}^d}\log G_\delta(x-y)\rho(r,dy) \leqslant \log\rho_\delta(r,x) = \log\int_{\mathbb{R}^d}G_\delta(x-y)\rho(r,dy) \leqslant \log\|G_\delta\|_\infty,$$

implying

$$\left|\int_{\mathbb{R}^d} \rho_\delta(s,x)\log\rho_\delta(s,x)\,dx\right| < \infty.$$

Combined with $\int_{\mathbb{R}^d}\Psi\,d\rho(s) < \infty$, $S(\rho_\delta(s)) < \infty$. Therefore, by Lemma 2.6, $\rho_\delta \in AC^2((s,T);\mathcal{P}_2(\mathbb{R}^d))$ and estimate (2.8) holds with ρ replaced by ρ_δ

$$\int_s^t I\big(\rho_\delta(r)\big)\,dr < \infty.$$

We now invoke Theorem 8.4.7 of [2], for each $0 < s < t < \infty$,

$$\begin{aligned}
&\frac{1}{2}d^2\big(\rho_\delta(t),\gamma\big) - \frac{1}{2}d^2\big(\rho_\delta(s),\gamma\big) \\
&\quad = \int_s^t \int_{\mathbb{R}^d} -\nabla p_{\rho_\delta(r),\gamma}(x)\left(\nu\frac{\nabla\rho_\delta}{\rho_\delta} + u_\delta - v_\delta\right)\rho_\delta(r,dx)\,dr \\
&\quad \leqslant -\nu\int_s^t\int_{\mathbb{R}^d} \nabla p_{\rho_\delta(r),\gamma}\left(\frac{\nabla\rho_\delta}{\rho_\delta} + \nabla\Psi\right)d\rho_\delta(r)\,dr \\
&\qquad + \nu\int_s^t\int_{\mathbb{R}^d}\nabla p_{\rho_\delta(r),\gamma}(\nabla\Psi - u_\delta)\,d\rho_\delta(r)\,dr + \int_s^t \big\|m(r)\big\|_{-1,\rho(r)}\,d\big(\rho_\delta(r),\gamma\big)\,dr \\
&\quad \leqslant \int_s^t\left(\nu S(\gamma) - \nu S\big(\rho_\delta(r)\big) - \frac{\lambda_\Psi}{2}\nu d^2\big(\rho_\delta(r),\gamma\big)\right)dr + \nu\int_s^t\int_{\mathbb{R}^d}\nabla p_{\rho_\delta(r),\gamma}(\nabla\Psi - u_\delta)\,d\rho_\delta(r)\,dr \\
&\qquad + \int_s^t\big\|m(r)\big\|_{-1,\rho(r)}\,d\big(\rho_\delta(r),\gamma\big)\,dr,
\end{aligned}$$

where first inequality follows by Lemma 8.1.9 of [2] and the second one follows from Theorem D.50 on p. 397 of [14] (see Lemma 5.2). In the above,

$$p_{\rho_\delta(r),\gamma}(x) := \frac{|x|^2}{2} - \varphi_{\rho_\delta(r),\gamma}(x)$$

with $\varphi = \varphi_{\rho_\delta(r),\gamma}$ a convex function such that $\nabla\varphi$ pushes forward the measure $\rho_\delta(r)$ to γ (i.e. $\nabla\varphi_{\#}\rho_\delta(r) = \gamma$). $p_{\rho_\delta(r),\gamma}$ is the difference between two convex functions defining Brenier's optimal transport map in Theorem D.25 (Appendix D) of [14]. See also Theorem 6.2.4 on p. 140 and Section 6.2.3 of [2].

We now pass $\delta \to 0+$. Note that, by (1.50),

$$\begin{aligned}
&\lim_{\delta\to 0+}\int_s^t\int_{\mathbb{R}^d}|\nabla\Psi - u_\delta|^2\,d\rho_\delta(r)\,dr \\
&\quad = \lim_{\delta\to 0+}\int_s^t\left(\|\nabla\Psi\|^2_{L^2_{\rho_\delta(r)}} + \|u_\delta\|^2_{L^2_{\rho_\delta(r)}} - 2\int_{\mathbb{R}^d} G_\delta * \nabla\Psi\cdot\nabla\Psi\,d\rho(r)\right)dr = 0;
\end{aligned}$$

(2.11) follows from lower semicontinuity of S.

In the second part of the proof, we extend the result allowing $\int_{\mathbb{R}^d}\Psi\,d\rho(s) = +\infty$.

We consider (2.5) and its approximation

$$X_k(t) = X_k(s) + \int_s^t \big(-\nu\nabla\Psi\big(X_k(r)\big) + U(r)\big)\,dr + \sqrt{2\nu}\big(W(t) - W(s)\big),$$

where $X_k(s) := X(s) \wedge k \vee (-k)$. Then by Ito's formula and by the assumption $D^2\Psi \geqslant \lambda_\Psi I$,

$$\begin{aligned}\frac{1}{2}\big|X_k(t) - X(t)\big|^2 &= \frac{1}{2}\big|X_k(s) - X(s)\big|^2 - \nu\int_s^t \big(X_k(r) - X(r)\big)\big(\nabla\Psi\big(X_k(r)\big) - \nabla\Psi\big(X(r)\big)\big)\,dr \\ &\leqslant \frac{1}{2}\big|X_k(s) - X(s)\big|^2 - \nu\lambda_\Psi\int_s^t \big|X_k(r) - X(r)\big|^2\,dr.\end{aligned}$$

Therefore, $\lim_{k\to\infty} X_k = X$ almost surely as $C([s,T];\mathbb{R}^d)$ valued random variables and

$$\sup_{s\leqslant r\leqslant T} E\big[\big|X_k(r) - X(r)\big|^2\big] \leqslant C_{s,T} E\big[X_k(s) - X(s)|^2\big].$$

Denote

$$\rho_k(r,dx) := P\big(X_k(r) \in dx\big), \quad r \in [s,T],$$

Note that, from (2.4), $\sup_{0\leqslant t\leqslant T} E[|X(t)|^2] \leqslant C(1 + E[|X(s)|^2]) < \infty$. Then

$$\lim_{k\to\infty}\sup_{s\leqslant r\leqslant T} d\big(\rho_k(r),\rho(r)\big) \leqslant \lim_{k\to\infty}\sup_{s\leqslant r\leqslant T} E\big[\big|X_k(r) - X(r)\big|^2\big] = 0.$$

Furthermore, by Ito's formula

$$\partial_r\rho_k = \nu\big(\Delta\rho_k + \operatorname{div}(\rho_k\nabla\Psi)\big) + m_k,$$

where m_k is defined distributionally as

$$\langle m_k, f\rangle := E\big[U(r)\nabla f\big(X_k(r)\big)\big], \quad \forall f \in C_c^\infty\big(\mathbb{R}^d\big).$$

It follows

$$\big|\langle m_k, f\rangle\big| \leqslant E^{1/2}\big[\big|U(r)\big|^2\big]E^{1/2}\big[\big|\nabla f\big(X_k(r)\big)\big|^2\big] \leqslant \|m\|_{-1,\rho(r)}\sqrt{\int_{\mathbb{R}^d}|\nabla f|^2\,d\rho_k}, \quad f \in C_c^\infty\big(\mathbb{R}^d\big).$$

Consequently,

$$\big\|m_k(r)\big\|_{-1,\rho_k(r)} \leqslant \big\|m(r)\big\|_{-1,\rho(r)}, \quad \forall r \in [s,T].$$

Note, additionally, that $|X_k(s)| \leqslant k$, ρ_k satisfies the condition of the first part of the proof. Therefore

$$\begin{aligned}\frac{1}{2}d^2\big(\rho_k(t),\gamma\big) + \nu\int_s^t S\big(\rho_k(r)\big)\,dr &\leqslant \frac{1}{2}d^2\big(\rho_k(s),\gamma\big) + \nu S(\gamma)(t-s) \\ &\quad + \int_s^t\Big[-\nu\frac{\lambda_\Psi}{2}d^2\big(\rho_k(r),\gamma\big) + d\big(\rho_k(r),\gamma\big)\big\|m(r)\big\|_{-1,\rho(r)}\Big]\,dr.\end{aligned}$$

Passing $k \to \infty$ and using the lower semicontinuity of $d^2(\rho,\gamma)$ and $S(\rho)$ with respect to ρ in the weak convergence in probability measure topology, we conclude. □

Suppose that ρ has finite action, or equivalently finite kinetic energy (2.1). Taking $s = 0$ in the above lemma, we know that $S(\rho(r)) < \infty$ for $r > 0$ almost everywhere. Then by Lemma 2.5, $S(\rho(r)) < \infty$ for all $r > 0$.

By (9.98) in [14],

$$\operatorname{grad} S\big(\rho(t)\big) = -\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big) \in \mathcal{D}'\big(\mathbb{R}^d\big), \quad t > 0. \tag{2.12}$$

Combine Lemmas 2.6, 2.7, the conclusion of Lemma 2.1 follows.

2.2. Some properties of relative entropies

Taking $\nu = 0$ in (1.22), our definition of kinetic energy integral reduces to the conventional form, which has been used in continuum mechanics. The work of Benamou and Brenier [3] showed that minimizer of such time integral gives Wasserstein-2 metric. With $\nu > 0$, we lose the geodesic interpretation of the minimizing path even at a formal level. This is because that, the minimizer, even if it exists, is not time reversible. In Lemma 2.10, we identify the value of such minimizer using relative entropy, a time asymmetric concept. The proof of this lemma relies on an intrinsic connection with probability theory, of which a heuristic discussion can be found in Appendix B.

2.2.1. Relative entropy and minimizer of kinetic energy

We discuss a number of properties concerning the relative entropy defined by (1.27).

First, the following is a consequence of elementary approximation by truncation and mollification.

Lemma 2.8. *Let $P, Q \in \mathcal{P}(\mathbb{R}^m)$ and $S(P\|Q) < \infty$, then for each $\epsilon > 0$, there exists a $g_\epsilon \in C_c(\mathbb{R}^m)$ such that the probability measure P_ϵ defined by*

$$P_\epsilon(dx) := Z_\epsilon^{-1} e^{g_\epsilon(x)} Q(dx)$$

has the following properties:

(1) *the total variation norm $\|P - P_\epsilon\|_T < \epsilon$;*
(2) *$S(P_\epsilon\|Q) \leqslant \epsilon + S(P\|Q)$.*

Let $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ be given. Let $Q(x, dy)$ be a transition probability measure on $\mathbb{R}^d$. That is $Q(x;\cdot) \in \mathcal{P}(\mathbb{R}^d)$ for each $x \in \mathbb{R}^d$ and $Q(\cdot, A) \in B(\mathbb{R}^d)$ for each Borel set $A \subset \mathbb{R}^d$. We denote

$$(Q \otimes \rho)(dx, dy) := Q(x, dy)\rho(dx) \in \mathcal{P}\big(\mathbb{R}^d \times \mathbb{R}^d\big), \quad \rho \in \mathcal{P}\big(\mathbb{R}^d\big).$$

For each $\pi \in \Pi(\rho_0, \rho_1)$ (see (1.8) for definition), we denote its transition probability by $\pi(x, dy)$ as well

$$\pi(dx, dy) = \pi(x, dy)\rho_0(dx) = (\pi \otimes \rho_0)(dx, dy).$$

Let $P(t) = P_t$ be the solution semigroup (i.e. $\rho(t) = P(t)\rho(0)$) of the Fokker–Planck equation

$$\partial_t \rho = \Delta\rho + \operatorname{div}(\rho\nabla\Psi). \tag{2.13}$$

We define

$$D(\rho_1\|\rho_0; t) := \inf_{\pi \in \Pi(\rho_0,\rho_1)} S(\pi\|P_{\nu t} \otimes \rho_0), \quad \forall \rho_0, \rho_1 \in \mathcal{P}_2\big(\mathbb{R}^d\big). \tag{2.14}$$

Then the following property holds

$$S(\pi\|P_{\nu T} \otimes \rho_0) = \int_{\mathbb{R}^d} S\big(\pi(x,\cdot)\big\| P_{\nu T}(x,\cdot)\big)\rho_0(dx).$$

A stochastic connection will be useful in relating the above defined D and the action functional.

Let $W_1, W_2 \ldots$ be a countable family of $\mathbb{R}^d$-valued independent standard Brownian motions. We define a collection of independent Markov processes $\{X_i(\cdot)\colon i = 1, 2, \ldots\}$ satisfying

(1) $\{X_i(0)\colon i = 1, 2, \ldots\}$ are independent identically distributed according to $P(X_i(0) \in dx) = \rho_0(dx)$;
(2)

$$dX_i(t) = \sqrt{2}dW_i(t) - \nabla\Psi\big(X_i(t)\big)\,dt, \quad i = 1, 2, \ldots. \tag{2.15}$$

The generator for X_i is

$$B := \Delta_\Psi = e^{\Psi} \operatorname{div}\big(e^{-\Psi}\nabla\big) = \Delta - \nabla\Psi \cdot \nabla.$$

Generator for a time rescaled version $X_i(\nu\cdot)$ is νB. We denote transition probability for X_i by $P_t(x, dy)$ its density $p(t; x, y)$ and the associated transition semigroup $P(t)$:

$$P_t(x, dy) := p(t; x, y)\, dy := P\big(X_i(t) \in dy \big| X_i(0) = x\big), \tag{2.16}$$

$$\rho(t) = \rho(t, dy) := P(t)\rho_0(y) = \int_{\mathbb{R}^d} p(t; x, y)\rho_0(dx). \tag{2.17}$$

Then P_t is the solution semigroup to (2.13).

Define stochastic empirical-measure-valued process

$$\mu_n(t, dx) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\nu t)}(dx). \tag{2.18}$$

Then by a straightforward modification of Theorem 13.37 of [14] (where only $\nu = 1$ was considered), or adaptations of [10] (where μ_n was studied with a different weak topology in state space $\mathcal{P}(\mathbb{R}^d)$), we have the following.

Lemma 2.9. *The process $\{\mu_n(\cdot)\colon n = 1, 2, \ldots\}$ given by* (2.18) *satisfies a large deviation principle in $C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ with good rate function $\nu^{-1}K_T[\cdot]$. That is,*

(1) *for each $\sigma \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ satisfying $\sigma(0) = \rho_0$, we have*

$$-\lim_{\epsilon\to 0+}\lim_{n\to\infty}\frac{1}{n}\log P\Big(\sup_{0\leqslant t\leqslant T} d\big(\mu_n(t), \sigma(t)\big) < \epsilon\Big) = \nu^{-1}K_T[\sigma];$$

(2) *the level sets $\{\rho \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d))\colon\ K_T[\rho] \leqslant C\}$ is compact in $C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ for all $C \in \mathbb{R}$.*

The next lemma will play key roles in several technical estimates in later sections.

Lemma 2.10. *For every $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$\inf\big\{K_T\big[\sigma(\cdot)\big]\colon\ \sigma(\cdot) \in \Gamma(\rho_0, \rho_1)\big\} = \nu D(\rho_1\|\rho_0; T). \tag{2.19}$$

Moreover, the minimum in (2.19) *is attained provided the right hand side of* (2.19) *is finite.*

Proof. The $\{X_1, X_2, \ldots\}$ are independent and identically distributed. For $T \geqslant 0$ fixed.

We define random measure

$$\pi_n(dx, dy) := \frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i(0), X_i(\nu T))}(dx, dy).$$

By the probabilistic Sanov's theorem (e.g. Theorem 6.2.10 of [9]), the $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$-valued random variables $\{\pi_n\colon n = 1, 2, \ldots\}$ satisfy a large deviation principle with good rate function $S(\pi \,\|\, P_{\nu T} \otimes \rho_0)$: for every $\pi \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$,

$$-\lim_{\epsilon\to 0+}\lim_{n\to\infty}\frac{1}{n}\log P\big(d(\pi_n, \pi) < \epsilon\big) = S(\pi \,\|\, P_{\nu T} \otimes \rho_0).$$

By a version of the contraction principle (e.g. Theorem 4.2.1 of [9]), consequently

$$-\lim_{\epsilon\to 0+}\lim_{n\to\infty}\frac{1}{n}\log P\big(d\big(\mu_n(0), \rho_0\big) + d\big(\mu_n(T), \rho_1\big) < \epsilon\big) = \inf_{\pi\in\Pi(\rho_0,\rho_1)} S(\pi \,\|\, P_{\nu T} \otimes \rho_0).$$

On the other hand, by another application of the contraction principle to the trajectory space level large deviation result in Lemma 2.9,

$$\lim_{\epsilon\to 0+}\lim_{n\to\infty}\frac{1}{n}\log P\big(d\big(\mu_n(0),\rho_0\big)+d\big(\mu_n(T),\rho_1\big)<\epsilon\big)=-\inf\big\{\nu^{-1}K_T\big[\sigma(\cdot)\big]:\ \sigma(0)=\rho_0,\sigma(T)=\rho_1\big\},$$

giving (2.19).

The attainment of minimum follows from the compact level set property, given in Lemma 2.9, for the functional K_T. □

By Lemma 2.10, there exists a path $\sigma(\cdot)\in\Gamma_T(\rho_0,\rho_1)$ with finite kinetic energy $K_T[\sigma(\cdot)]<\infty$ if and only if $D(\rho_1\|\rho_0;T)<\infty$. With the above estimates, we have a reachability type result for moving mass around with finite kinetic energy K_T. In particular, we have

$$\nu D\big(\rho(t)\big\|\rho_0;t\big)\leqslant K_t[\rho]\leqslant K_T[\rho]=\frac{1}{2}\int_0^T\big\|\dot\rho-\nu\big(\Delta\rho+\operatorname{div}(\rho\nabla\Psi)\big)\big\|^2_{-1,\rho}\,ds, \tag{2.20}$$

and

$$D(P_{\nu t}\rho_0\|\rho_0;t)=0,$$

for any $t\in[0,T]$ and any path $\rho(\cdot)\in C([0,T];\mathcal{P}_2(\mathbb{R}^d))$ with $\rho(0)=\rho_0$.

In view of Lemma 2.1 (taking $F=0$), at least when $S(\rho_0)<\infty$, (2.19) can be improved into

$$\nu D(\rho_1\|\rho_0;T)=\inf\left\{K_T\big[\sigma(\cdot)\big]:\ \sigma\in\Gamma(\rho_0,\rho_1),\int_0^T\big(\|\dot\sigma\|^2_{-1,\sigma}+I(\sigma)\big)\,ds<\infty\right\}. \tag{2.21}$$

Lemma 2.11. *Assume that $S(\rho_0)+S(\rho_1)<\infty$, then*

$$D(\rho_0\|\rho_1;T)=D(\rho_1\|\rho_0;T)-2\big(S(\rho_1)-S(\rho_0)\big).$$

Proof. Let $\sigma(\cdot)$ be as in the right hand side of (2.21) and we define $\hat\sigma(t):=\sigma(T-t)$. Then

$$\begin{aligned}K_T\big[\hat\sigma(\cdot)\big]&=\frac{1}{2}\int_0^T\big\|\dot{\hat\sigma}+\nu\operatorname{grad}S(\hat\sigma)\big\|^2_{-1,\hat\sigma}\,ds\\&=\frac{1}{2}\int_0^T\big\|\dot\sigma-\nu\operatorname{grad}S(\sigma)\big\|^2_{-1,\sigma}\,ds=K_T\big[\sigma(\cdot)\big]-2\nu\big(S(\rho_1)-S(\rho_0)\big),\end{aligned}$$

where the last equality follows from (2.10) in Lemma 2.6 (applied to $\sigma(\cdot)$).

Hence the conclusion holds. □

Notice that $D(\cdot\|\cdot;t)$ is not a distance. In particular, there is no triangle inequality. However, the following holds.

Lemma 2.12. *For any $\rho,\gamma,\sigma\in\mathcal{P}_2(\mathbb{R}^d)$, and $r+s=t$ with $r,s\geqslant 0$,*

$$D(\gamma\|\rho;t)\leqslant D(\sigma\|\rho;r)+D(\gamma\|\sigma;s).$$

Proof. We assume that the right hand side of the inequality is finite, otherwise the inequality holds trivially. By Lemma 2.10, there exists $\sigma_1(\cdot)\in C([0,r];\mathcal{P}_2(\mathbb{R}^d))$ and $\sigma_2(\cdot)\in C([0,s];\mathcal{P}_2(\mathbb{R}^d))$ such that

$$\sigma_1(0)=\rho,\qquad\sigma_1(r)=\sigma,\qquad\nu D(\sigma\|\rho;r)=\int_0^r T\big(\sigma_1(u),\dot\sigma_1(u)\big)\,du$$

and

$$\sigma_2(0) = \sigma, \qquad \sigma_2(s) = \gamma, \qquad \nu D(\gamma \| \sigma; s) = \int_0^s T\big(\sigma_2(u), \dot{\sigma}_2(u)\big)\, du.$$

Pasting σ_1, σ_2 together to define a new trajectory $\sigma(\cdot) \in C([0,t]; \mathcal{P}_2(\mathbb{R}^d))$ with $\sigma(0) = \rho$ and $\sigma(t) = \gamma$:

$$\sigma(u) = \sigma_1(u)\mathbf{1}_{0 \leqslant u \leqslant r} + \sigma_2(u - r)\mathbf{1}_{r \leqslant u \leqslant t}.$$

Then

$$\begin{aligned} \nu D(\gamma \| \rho; t) &\leqslant \int_0^t T\big(\sigma(u), \dot{\sigma}(u)\big)\, du \\ &= \int_0^r T\big(\sigma_1(u), \dot{\sigma}_1(u)\big)\, du + \int_0^s T\big(\sigma_2(u), \dot{\sigma}_2(u)\big)\, du \\ &= \nu D(\sigma \| \rho; r) + \nu D(\gamma \| \sigma; s). \qquad \square \end{aligned}$$

The following result follows essentially from a re-parameterization of time variable in the definition of action integral $K_T[\cdot]$ in (1.22), which in turn defines D.

Lemma 2.13. *Assume that $t > 0$, then*

$$\lim_{s \to 0+} D(\gamma \| \rho_0; t - s) = D(\gamma \| \rho_0; t).$$

Proof. First, we show that

$$\liminf_{s \to 0^+} D(\gamma \| \rho_0; t - s) \geqslant D(\gamma \| \rho_0; t).$$

We only need to prove this when $\liminf_{s \to 0^+} D(\gamma \| \rho_0; t - s) < \infty$. By working with a subsequence if necessary, we can assume without loss of generality that $D(\gamma \| \rho_0; t - s) < \infty$ for all $0 < s < t$. For each $0 < s < t$, let $\pi_s \in \Pi(\rho_0, \gamma)$ be such that $D(\gamma \| \rho_0; t - s) = S(\pi_s \| P_{\nu(t-s)} \otimes \rho_0)$. Let π be a limit point of the relatively compact sequence $\{\pi_s\colon s > 0\}$ in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Then $\pi \in \Pi(\rho_0, \gamma)$ and by lower semicontinuity of the relative entropy function,

$$\liminf_{s \to 0+} D(\gamma \| \rho_0; t - s) = \liminf_{s \to 0+} S(\pi_s \| P_{\nu(t-s)} \otimes \rho_0) \geqslant S(\pi \| P_{\nu t} \otimes \rho_0) \geqslant D(\gamma \| \rho_0; t).$$

Next we show that

$$\limsup_{s \to 0^+} D(\gamma \| \rho_0; t - s) \leqslant D(\gamma \| \rho_0; t).$$

We only need to show that case when $D(\gamma \| \rho_0; t) < \infty$. By Lemma 2.10, there exists $\sigma(\cdot) \in \Gamma_t(\rho_0, \gamma)$ such that

$$\nu D(\gamma \| \rho_0; t) = K_t\big[\sigma(\cdot)\big] = \int_0^t \frac{1}{2} \big\| \dot{\sigma} - \nu\big(\Delta\sigma + \operatorname{div}(\sigma \nabla \Psi)\big)\big\|_{-1,\sigma}^2 \, d\tilde{r} < \infty.$$

Let $\delta > 0$ be small enough so that $\delta < t - s$. We construct a new path

$$\bar{\sigma}(r) := \sigma(r), \quad 0 \leqslant r \leqslant \delta, \qquad \bar{\sigma}(r) := \sigma\left(\delta + \frac{t - \delta}{(t - \delta) - s}(r - \delta)\right), \quad r \in [\delta, t - s].$$

Then $\bar{\sigma}(\cdot) \in \Gamma_{t-s}(\rho_0, \gamma)$. Consequently, by Lemma 2.1,

$$\begin{aligned}
\nu D(\gamma\|\rho_0; t-s) \leqslant K_{t-s}\big[\bar\sigma(\cdot)\big] &= K_\delta\big[\sigma(\cdot)\big] + \int_\delta^{t-s} \frac{1}{2}\big\|\dot{\bar\sigma} + \nu \operatorname{grad} S(\bar\sigma)\big\|_{-1,\bar\sigma}^2\, dr \\
&= K_\delta\big[\sigma(\cdot)\big] + \frac{t-\delta-s}{t-\delta}\int_\delta^t \frac{1}{2}\bigg\|\frac{t-\delta}{t-\delta-s}\dot\sigma + \nu \operatorname{grad} S(\sigma)\bigg\|_{-1,\sigma}^2\, d\tilde r \\
&= K_\delta\big[\sigma(\cdot)\big] + \frac{t-\delta}{t-\delta-s}\int_\delta^t \frac{1}{2}\|\dot\sigma\|_{-1,\sigma}^2\, d\tilde r + \frac{t-\delta-s}{t-\delta}\int_\delta^t \frac{\nu^2}{2} I(\sigma)\, d\tilde r \\
&\quad + \int_\delta^t \nu\big\langle \operatorname{grad} S(\sigma), \dot\sigma\big\rangle_{-1,\sigma}^2\, d\tilde r.
\end{aligned}$$

With the estimates in Lemma 2.1, taking limit $\limsup_{s\to 0^+}$ on both side of the inequality above, we have

$$\limsup_{s\to 0^+} \nu D(\gamma\|\rho_0; t-s) \leqslant K_t\big[\sigma(\cdot)\big] = \nu D(\gamma\|\rho_0; t). \qquad \square$$

Lemma 2.14. *Suppose that $\gamma \in \mathcal{P}_2(\mathbb{R}^d)$ satisfies $S(\gamma) < \infty$. Then if $\rho_n \to \rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$, we have*

$$\lim_{n\to+\infty} D(\gamma\|\rho_n; t) = D(\gamma\|\rho_0; t)$$

for every $t > 0$.

Proof. Let $\pi_n \in \Pi(\rho_n, \gamma)$. Then $\{\pi_n\colon n = 1, 2, \ldots\}$ is relatively compact with every limiting point satisfying $\pi_0 \in \Pi(\rho_0, \gamma)$. By variational representation of relative entropy function, along subsequence of the limiting point,

$$\liminf_{n\to\infty} S(\pi_n\|P_{\nu t}\otimes\rho_n) \geqslant S(\pi_0\|P_{\nu t}\otimes\rho_0) \geqslant D(\gamma\|\rho_0; t).$$

By the arbitrariness of π_n, $\liminf_{n\to\infty} D(\gamma\|\rho_n; t) \geqslant D(\gamma\|\rho_0; t)$.

Next, we show that the other direction of the inequality also holds. Without loss of generality, we assume that $D(\gamma\|\rho_0; t) < \infty$. Let $0 < s < t$ and let $\pi_0(dx, dy) := \pi_0(x, dy)\rho_0(dx)$ be such that $D(\gamma\|\rho_0; t-s) = S(\pi_0\|P_{\nu(t-s)}\otimes\rho_0)$. The existence of such π_0 is guaranteed by the usual compactness-lower semicontinuity argument. We approximate such π_0 next using Lemma 2.8. We will find a function $h_\epsilon \in C_b(\mathbb{R}^d\times\mathbb{R}^d)$ and define the following quantities

$$\begin{aligned}
\pi_\epsilon(dx, dy) &:= Z_\epsilon^{-1} e^{h_\epsilon(x,y)}(P_{\nu(t-s)}\otimes\rho_0)(dx, dy), \\
\pi_{n,\epsilon}(dx, dy) &:= Z_{n,\epsilon}^{-1} e^{h_\epsilon(x,y)}(P_{\nu(t-s)}\otimes\rho_n)(dx, dy), \\
\gamma_{n,\epsilon}(dy) &:= \pi_{n,\epsilon}\big(\mathbb{R}^d, dy\big).
\end{aligned}$$

By Lemma 2.8, we can choose h_ϵ so that

$$D(\gamma\|\rho_0; t-s) = S(\pi_0\|P_{\nu(t-s)}\otimes\rho_0) \geqslant S(\pi_\epsilon\|P_{\nu(t-s)}\otimes\rho_0) - \epsilon. \tag{2.22}$$

Therefore,

$$D(\gamma_{n,\epsilon}\|\rho_n; t-s) \leqslant S(\pi_{n,\epsilon}\|P_{\nu(t-s)}\otimes\rho_n) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \big(h_\epsilon e^{h_\epsilon}\big)(P_{\nu(t-s)}\otimes\rho_n)(dx, dy) - \log Z_{n,\epsilon}.$$

Consequently

$$\begin{aligned}
\limsup_{n\to\infty} D(\gamma_{n,\epsilon}\|\rho_n; t-s) &\leqslant \int_{\mathbb{R}^d\times\mathbb{R}^d} \big(h_\epsilon e^{h_\epsilon}\big)(P_{\nu(t-s)}\otimes\rho_0)(dx, dy) - \log Z_\epsilon \\
&= S(\pi_\epsilon\|P_{\nu(t-s)}\otimes\rho_0) \leqslant \epsilon + D(\gamma\|\rho_0; t-s),
\end{aligned}$$

where the last inequality follows from (2.22).

By "the quasi-triangle inequality" in Lemma 2.12, for every $0 < s < t$,

$$D(\gamma \| \rho_n; t) \leqslant D(\gamma_{n,\epsilon} \| \rho_n; t - s) + D(\gamma \| \gamma_{n,\epsilon}; s).$$

Note that we suppress the dependence of $\gamma_{n,\epsilon} = \gamma_{n,\epsilon,s}$ on $s > 0$. For each $s > 0$ fixed, by Lemma 2.8, $\lim_{\epsilon\to 0+} \lim_{n\to\infty} \gamma_{n,\epsilon} = \gamma$ in narrow convergence topology in $\mathcal{P}(\mathbb{R}^d)$. We claim that (Lemma 2.16)

$$\lim_{s\to 0+} \lim_{\epsilon\to 0+} \lim_{n\to\infty} D(\gamma \| \gamma_{n,\epsilon}; s) = 0.$$

Take limits $\lim_{s\to 0+} \lim_{\epsilon\to 0+} \lim_{n\to\infty}$ on both sides, therefore

$$\limsup_{n\to\infty} D(\gamma \| \rho_n; t) \leqslant \limsup_{s\to 0+} D(\gamma \| \rho_0; t - s) + \lim_{s\to 0+} \lim_{\epsilon\to 0+} \lim_{n\to\infty} D(\gamma \| \gamma_{n,\epsilon}; s) \leqslant D(\gamma \| \rho_0; t),$$

where the last inequality above follows from Lemmas 2.13. □

Lemma 2.15. *Let $\gamma_1, \gamma_{0,s,n} \in \mathcal{P}_2(\mathbb{R}^d)$ for each $n = 1, 2, \dots$ and $s > 0$. Suppose that*

$$\sup_{n=1,2,\dots,s>0} S(\gamma_{0,s,n}) + S(\gamma_1) + I(\gamma_{0,s,n}) + I(\gamma_1) < \infty, \tag{2.23}$$

and that $\lim_{n\to\infty} d(\gamma_{0,s,n}, \gamma_1) = 0$ for each $s > 0$ fixed. Then

$$\lim_{s\to 0+} \lim_{n\to\infty} D(\gamma_1 \| \gamma_{0,s,n}; 2s) = 0.$$

Proof. To avoid long notations, we write $\gamma_{0,n} := \gamma_{0,s,n}$ while keeping in mind that the sequence depends on s.

Let $G_\tau(x) := (2\pi\tau)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2\tau}}$. Then by Lemma 2.12,

$$D(\gamma_1 \| \gamma_{0,n}; 2s) \leqslant D(\gamma_1 \| G_{\nu s} * \gamma_1; s) + D(G_{\nu s} * \gamma_1 \| \gamma_{0,n}; s).$$

First, we estimate the last term on the right hand side of the last equality. Let convex function $\varphi := \varphi_{\gamma_{0,n},\gamma_1}$ give the optimal transport map $\nabla\varphi_{\#}\gamma_{0,n} = \gamma_1$ (see notations in Example 1.4). Let X_0 be an $\mathbb{R}^d$-valued random variable with law $\gamma_{0,n}$ and $W(\cdot)$ be a standard Brownian motion independent of X_0, $W(0) = 0$. Fix $s > 0$, we consider stochastic differential equation

$$X(t) := X_0 + t\big(\nabla\varphi(X_0) - X_0\big) + \sqrt{2s\nu}\big(W(t) - W(0)\big), \quad t \in [0, 1].$$

Denote $\gamma(t; dx) := P(X(t) \in dx)$ and define Schwartz distribution m by

$$\big\langle m(t), f\big\rangle := E\big[\big(\nabla\varphi(X_0) - X_0\big)\nabla f\big(X(t)\big)\big], \quad \forall f \in C_c^\infty\big(\mathbb{R}^d\big).$$

Then

$$\begin{aligned}\big\|m(t)\big\|^2_{-1,\gamma(t)} &= \sup_{f\in C_c^\infty(\mathbb{R}^d)} \big\{2\big\langle m(t), f\big\rangle - E\big[\big|\nabla f\big(X(t)\big)\big|^2\big]\big\} \\ &\leqslant E\big[\big|\nabla\varphi(X_0) - X_0\big|^2\big] = d^2(\gamma_{0,n}, \gamma_1).\end{aligned}$$

Note that

$$\partial_t \gamma = s\nu\Delta\gamma + m.$$

Since $X_0 + t(\nabla\varphi(X_0) - X_0)|_{t=1} = \nabla\varphi(X_0)$ has probability law γ_1, therefore $\gamma(1) = G_{\nu s} * \gamma_1$. Moreover, $\gamma(0) = \gamma_{0,n}$. Hence through (2.19), by a reparametrization of time,

$$\begin{aligned}&D(G_{\nu s} * \gamma_1 \| \gamma_{0,n}; s) \\ &\quad = \frac{1}{s\nu} \inf\left\{ \frac{1}{2}\int_0^1 \big\|\dot\rho - s\nu\big(\Delta\rho + \operatorname{div}(\rho\nabla\Psi)\big)\big\|^2_{-1,\rho(t)}\, dt\colon\ \rho(\cdot) \in \varGamma_{T=1}(\gamma_{0,n}, G_{\nu s} * \gamma_1)\right\}\end{aligned}$$

$$\leqslant \frac{1}{s\nu}\int_0^1 \|\dot\gamma - s\nu\Delta\gamma\|_{-1,\gamma}^2\,dt + s\nu\int_0^1\int_{x\in\mathbb{R}^d} |\nabla\Psi|^2\gamma(t,dx)\,dt$$

$$\leqslant \frac{1}{s\nu}d^2(\gamma_1,\gamma_{0,n}) + s\nu\int_0^1 E\big[\big|\nabla\Psi\big(X(t)\big)\big|^2\big]\,dt.$$

Denote $Y_0 = \nabla\varphi(X_0)$. By (1.50),

$$\begin{aligned}&E\big[\big|\nabla\Psi\big(X(t)\big) - \nabla\Psi\big((1-t)X_0+tY_0\big)\big|^2\big]\\&\quad\leqslant E\big[\omega\big(\sqrt{2s\nu}W(t)\big)\big]E\big[(1+\big|\nabla\Psi\big((1-t)X_0+tY_0\big)\big|^2+\Psi\big((1-t)X_0+tY_0\big)\big].\end{aligned}$$

Furthermore, by semi-convexity of ψ and Ψ, there exists $C\in\mathbb{R}$ such that

$$E\big[\psi\big((1-t)X_0+tY_0\big)\big]\leqslant(1-t)E\big[\psi(X_0)\big]+tE\big[\psi(Y_0)\big]+Ct(1-t)E\big[|X_0-Y_0|^2\big];$$
$$E\big[\Psi\big((1-t)X_0+tY_0\big)\big]\leqslant(1-t)E\big[\Psi(X_0)\big]+tE\big[\Psi(Y_0)\big]+Ct(1-t)E\big[|X_0-Y_0|^2\big].$$

Moreover, by (2.23) and (1.51),

$$\begin{aligned}&\sup_{n,s}E\big[\big|\nabla\Psi(Y_0)\big|^2+\big|\nabla\Psi(X_0)\big|^2+\Psi(X_0)+\Psi(Y_0)\big]\\&\quad=\sup_{n,s}\int_{\mathbb{R}^d}|\nabla\Psi|^2\,d\gamma_{0,n}+\int_{\mathbb{R}^d}|\nabla\Psi|^2d\gamma_1+\int_{\mathbb{R}^d}\Psi\,d\gamma_{0,s,n}+\int_{\mathbb{R}^d}\Psi\,d\gamma_1<\infty.\end{aligned}$$

Combine all the above together, $\lim_{s\to0+}\lim_{n\to\infty}D(G_{\nu s}*\gamma_1\|\gamma_{0,s,n};s)=0$.

Next, we show that $\lim_{s\to0+}D(\gamma_1\|G_{\nu s}*\gamma_1;s)=0$ hence conclude the lemma. Denote $\rho(t):=G_{\nu t}*\gamma_1$, then $\partial_t\rho=\nu\Delta\rho$. Let $\sigma(t)=\rho(s-t)$ for $0\leqslant t\leqslant s$. Then $\sigma(0)=G_{\nu s}*\gamma_1$ and $\sigma(s)=\gamma_1$, $\dot\sigma=-\nu\Delta\sigma$. Consequently

$$\begin{aligned}D(\gamma_1\|G_{\nu s}*\gamma_1;s)&\leqslant\frac{1}{2\nu}\int_0^s\big\|\dot\sigma-\nu\big(\Delta\sigma+\operatorname{div}(\sigma\nabla\Psi)\big)\big\|_{-1,\sigma}^2\,dt\\&=2\nu\int_0^s\left\|\Delta\sigma+\operatorname{div}\left(\sigma\nabla\frac{\Psi}{2}\right)\right\|_{-1,\sigma}^2\,dt\\&\leqslant4\nu\int_0^s\int_{\mathbb{R}^d}\frac{|\nabla_x\rho(t,x)|^2}{\rho(t,x)}\,dx\,dt+\nu\int_0^s\int_{\mathbb{R}^d}|\nabla\Psi|^2\rho(t,dx)\,dt.\end{aligned}$$

By variational representation of the Fisher information and Jensen's inequality (e.g. Lemma 8.1.10 of [2]),

$$\int_{\mathbb{R}^d}\frac{|\nabla\rho(t,x)|^2}{\rho(t,x)}\,dx=\int\frac{|\nabla G_{\nu t}*\gamma_1|^2}{G_{\nu t}*\gamma_1}\,dx\leqslant\int\frac{|\nabla\gamma_1|^2}{\gamma_1}\,dx.$$

Note that $\rho(t,x)=P(X+\sqrt{2\nu t}Z\in dx)$ where random variable X has law γ_1 and Z is a standard normal random variable. By argument similar to earlier, using (1.50)

$$\begin{aligned}\int\big|\nabla\Psi(x)\big|^2\rho(t,dx)&=E\big[\big|\nabla\Psi(X+\sqrt{2\nu t}Z)\big|^2\big]\leqslant C\big(E\big[\big|\nabla\Psi(X)\big|^2\big]+E\big[\Psi(X)\big]+1\big)\\&=C\left(1+\int|\nabla\Psi|^2\,d\gamma_1+\int\Psi\,d\gamma_1\right).\end{aligned}$$

Therefore

$$\lim_{s\to0+}D(\gamma_1\|G_{\nu s}*\gamma_1;s)=0.\qquad\square$$

Lemma 2.16. *Assume that $S(\gamma_0) < \infty$, and that for each $s > 0$, $\lim_{n\to\infty} \gamma_{n,s} = \gamma_0$ in the weak convergence of probability measure (i.e. the narrow convergence) topology. Then*

$$\lim_{s\to 0+} \lim_{n\to\infty} D(\gamma_0 \| \gamma_{n,s}; s) = 0.$$

Proof. We will write $\gamma_n := \gamma_{n,s}$ to simplify writing as there will be further subindex introduced.

By Lemma 2.12, for every $0 < 2r < s$,

$$\begin{aligned} D(\gamma_0\|\gamma_n; s) &\leqslant D(\gamma_0\|P_{\nu r}\gamma_0; r) + D(P_{\nu r}\gamma_0\|P_{\nu r}\gamma_n; s-2r) + D(P_{\nu r}\gamma_n\|\gamma_n; r) \\ &= 0 - 2\big(S(P_{\nu r}\gamma_0) - S(\gamma_0)\big) + D(P_{\nu r}\gamma_0\|P_{\nu r}\gamma_n; s-2r) + 0, \end{aligned}$$

where we used Lemma 2.11 in the last step.

By the estimate in Lemma 2.15, for each $s > 0$ fixed,

$$\lim_{r\to\frac{s}{2}} \lim_{n\to\infty} D(P_{\nu r}\gamma_0\|P_{\nu r}\gamma_n; s-2r) = 0.$$

Moreover, by (2.10) in Lemma 2.6, denote $\gamma(t) = P_{\nu t}\gamma_0$, then

$$\lim_{s\to 0+} \lim_{r\to\frac{s}{2}} \nu\big(S(P_{\nu r}\gamma_0) - S(\gamma_0)\big) = -\lim_{s\to 0+} \lim_{r\to\frac{s}{2}} \int_0^r \left(\frac{1}{2}\|\dot\gamma\|_{-1,\gamma}^2 + \frac{\nu^2}{2} I(\gamma)\right) dt = 0.$$

Consequently, we conclude that

$$\begin{aligned} \lim_{s\to 0+} \lim_{n\to\infty} D(\gamma_0\|\gamma_n; s) &\leqslant \lim_{s\to 0+} \lim_{r\to\frac{s}{2}} (-2)\big(S(P_{\nu r}\gamma_0) - S(\gamma_0)\big) \\ &\quad + \lim_{s\to 0+} \lim_{r\to\frac{s}{2}} \lim_{n\to\infty} D(P_{\nu r}\gamma_0\|P_{\nu r}\gamma_n; s-2r) = 0. \qquad \square \end{aligned}$$

Lemma 2.17. *Suppose $\rho_n \to \rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$, and let $\sigma(\cdot) \in C([0,T]; \mathcal{P}_2(\mathbb{R}^d))$ satisfy* (2.1). *Then*

$$\lim_{n\to\infty} D\big(\sigma(\tau)\|\rho_n; \tau\big) = D\big(\sigma(\tau)\|\rho_0; \tau\big)$$

for every $\tau \in (0, T)$.

Proof. Since $\sigma(\cdot)$ satisfies (2.1), we obtain from Lemma 2.1 that $S(\sigma(\tau)) < \infty$ for each $\tau > 0$. The result then follows from Lemma 2.14. $\square$

2.3. Internal energy/pressure function estimate

We provide some estimates regarding internal energy and pressure function which will be useful in later analysis.

Lemma 2.18. *Condition* 1.5 *implies Condition* 1.6. *That is, there exists a nondecreasing sub-linear function $\hat\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that $|V(\rho)| \leqslant \hat\zeta(I(\rho))$. Moreover, V is continuous in the weak convergence of probability measure topology in finite level sets of I.*

Proof. The proof of (1.55) consists of three parts:

First, we show that for $\beta \geqslant 1$,

$$\int_{\mathbb{R}^d} \rho^\beta(x)\,dx \leqslant C\left(1 + \int_{\mathbb{R}^d} |\nabla\sqrt{\rho}|^2\,dx\right)^{r_0} \leqslant C\big(1 + I(\rho)\big)^{r_0} \tag{2.24}$$

for some $r_0 < 1$. Let $\rho(dx) = \rho(x)\,dx \in \mathcal{P}_2(\mathbb{R}^d)$. Take $0 \leqslant \delta \leqslant 1$ and $p \geqslant 1$ satisfying $\delta\beta p = 1$, then

$$\int_{\mathbb{R}^d} \rho^\beta(x)\,dx = \int_{\mathbb{R}^d} \rho^{\beta\delta}\rho^{\beta(1-\delta)}\,dx \leqslant \Big(\int_{\mathbb{R}^d} \rho^{\beta\delta p}\,dx\Big)^{\frac{1}{p}} \Big(\int_{\mathbb{R}^d} \rho^{\beta\frac{(1-\delta)p}{p-1}}\,dx\Big)^{\frac{p-1}{p}}$$
$$= \Big(\int_{\mathbb{R}^d} \rho^{\frac{\beta p-1}{p-1}}\,dx\Big)^{\frac{p-1}{p}} = \|\sqrt{\rho}\|_{\frac{2(\beta p-1)}{p-1}}^{\frac{2(\beta p-1)}{p}}.$$

We recall the usual Sobolev embedding theorem (e.g. Theorem 4.12 in Adams and Fournier [1]) next. In dimension one and two $d=1,2$,

$$\|f\|_{L^q(\mathbb{R}^d)} \leqslant C\|f\|_{H^1(\mathbb{R}^d)} \tag{2.25}$$

for any $q \in [2,\infty)$. Therefore provided $\beta < 2$, we can find a p satisfying the above requirement and (2.24) holds with $r_0 = \beta - 1/p < 1$. In dimension three and beyond $d \geqslant 3$, Sobolev inequality (2.25) holds with $q \in [2, \frac{2d}{d-2}]$. When $\beta \in (1, \frac{d+2}{d})$, $d-(d-2)\beta > 0$, and

$$0 < \frac{2}{d-(d-2)\beta} < \frac{1}{\beta-1} < \infty.$$

Hence we can take a $p > 1$ satisfying $\frac{2}{d-(d-2)\beta} < p$ and $p < \frac{1}{\beta-1}$ at the same time. The first inequality implies $\frac{\beta p-1}{p-1} < \frac{d}{d-2}$, and the second implies $\frac{\beta p-1}{p} < 1$. Consequently (2.24) follows again. In view of the above estimates and (1.54),

$$\int_{\{x\in\mathbb{R}^d:\ |\rho(x)|\geqslant 1\}} \big|F\big(\rho(x)\big)\big|\,dx \leqslant C\big(1+I(\rho)\big)^{r_0}.$$

Second, from (1.53), we have $|F(s)| \leqslant Cs^\alpha$, $s \in (0,1)$. Since $\alpha < 1$ and $2\alpha/(1-\alpha) > d$,

$$\begin{aligned}\int_{\{x\in\mathbb{R}^d:\ |\rho(x)|<1\}} \big|F\big(\rho(x)\big)\big|\,dx &\leqslant C\int_{\mathbb{R}^d} \rho^\alpha\,dx \\ &= \int_{\mathbb{R}^d} \rho^\alpha(x)\big(1+|x|\big)^{2\alpha}\big(1+|x|\big)^{-2\alpha}\,dx \\ &\leqslant \Big(\int_{\mathbb{R}^d}\big(1+|x|^2\big)\rho(x)\,dx\Big)^\alpha \Big(\int_{\mathbb{R}^d}\big(1+|x|^2\big)^{-\frac{\alpha}{1-\alpha}}\,dx\Big)^{1-\alpha} \\ &\leqslant C\Big(1+\int_{\mathbb{R}^d}|x|^2\,d\rho\Big).\end{aligned}$$

Since ψ has super-quadratic growth, there exists sub-linear, concave function $\zeta : R_+ \mapsto \mathbb{R}$ with $|x|^2 \leqslant \zeta \circ \psi(x)$. Therefore

$$\int_{\{x\in\mathbb{R}^d:\ |\rho(x)|<1\}} \big|F\big(\rho(x)\big)\big|\,dx \leqslant \int_{\mathbb{R}^d} (\zeta\circ\psi)\,d\rho \leqslant \zeta\Big(\int_{\mathbb{R}^d}\psi\,d\rho\Big) \leqslant \zeta\big(I(\rho)\big),$$

where the last step follows from Jensen's inequality.

By assumptions regarding Φ and ϕ in Condition 1.5, they may have at most quadratic growth. Therefore, combined with the previous estimates for $\int_{\mathbb{R}^d} F(\rho(x))\,dx$, we have (1.55).

Finally, let $\rho_n \Rightarrow \rho$ in weak convergence of probability measure (i.e. narrow convergence) topology and suppose $\sup_n I(\rho_n) < \infty$. In particular $\sup_n \int_{\mathbb{R}^d} |\nabla\sqrt{\rho_n}|^2\,dx < \infty$, hence up to a subsequence $\sqrt{\rho_n} \to \sqrt{\rho}$ in L^2_{loc} norm topology by compact Sobolev embedding result. For details, see similar arguments in the proof of Lemma D.48 on p. 396 of [14]. This implies $F(\rho_n(x)) \to F(\rho(x))$ almost everywhere. Next, we show that the ζ can always be chosen so that

$$\int_{\mathbb{R}^d} \left|F\big(\rho(x)\big)\right|^{1+\epsilon} dx \leqslant \eta\big(I(\rho)\big) \leqslant C < \infty \tag{2.26}$$

for some $\epsilon > 0$. By uniform integrability, therefore

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} F\big(\rho_n(x)\big)\, dx = \int_{\mathbb{R}^d} F\big(\rho(x)\big)\, dx.$$

To see (2.26) holds, we observe that if F satisfies (1.53) and (1.54), $|F|^{1+\epsilon}$ also satisfies the conditions for some sufficiently small $\epsilon > 0$ with a slightly different choice of α, β (for instance, by $(1+\epsilon)\alpha$ and $(1+\epsilon)\beta$).

Since $\sup_n I(\rho_n) < \infty$ also implies that $\sup_n \int_{\mathbb{R}^d} \psi\, d\rho_n < \infty$. ψ has super-quadratic growth and Φ, ϕ have at most quadratic growth, therefore interaction energy W defined in (1.4) converges as well

$$\lim_{n\to+\infty} W(\rho_n) = W(\rho). \qquad \square$$

3. Action-minimizing paths in the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$

3.1. Existence of action minimizer

We assume that (1.7) holds and that extension of V to all probability measures satisfies $\|V \vee 0\|_\infty < \infty$. Then

$$A_T[\rho] = K_T[\rho] - \int_0^T V\big(\rho(s)\big)\, ds : C\big([0,T]; \mathcal{P}_2\big(\mathbb{R}^d\big)\big) \mapsto [-c, +\infty],$$

where $c = \|V \vee 0\|_\infty T$, is well defined. In some interesting cases,

$$\int_0^T V\big(\rho(s)\big)\, ds : \rho(\cdot) \in C\big([0,T]; \mathcal{P}_2\big(\mathbb{R}^d\big)\big) \mapsto [-\infty, c]$$

can be upper semicontinuous. For instance, take

$$F(r) = -c_1 r^\beta + c_2 r^\alpha + c_3 r \log r,$$

where $c_1 > 0$ and $1 \leqslant \alpha < \beta$, and $\beta \in (1, \frac{d+2}{d})$, $d \geqslant 3$ and $\beta \in (1,2)$ when $d = 1, 2$. We define

$$V(\rho) = \begin{cases} \int_{\mathbb{R}^d} F(\rho(x))\, dx + W(\rho) & \text{when } \rho(dx) = \rho(x)\, dx, \\ -\infty & \text{otherwise.} \end{cases}$$

Then by a result regarding convex integrals of measures (e.g. Lemma 9.4.4 of [2]) applied to the leading order term $-c_1 r^\beta$ of F, and since r^α and $r \log r$ can all be dominated by r^β when $r \geqslant 1$, we have that V is upper semicontinuous in the Wasserstein convergence topology (W is continuous in this topology).

Lemma 3.1. *Assume that $\|V \vee 0\|_\infty < \infty$ and Condition* 1.5 *hold, and that V is upper semicontinuous on $\mathcal{P}_2(\mathbb{R}^d)$. Let $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$. Then there exists a path $\rho(\cdot) \in \Gamma(\rho_0, \rho_1) \cap AC^2(0, T; \mathcal{P}(\mathbb{R}^d))$ such that*

$$A_T\big[\rho(\cdot)\big] = \inf\big\{A_T\big[\sigma(\cdot)\big]:\ \sigma(\cdot) \in \Gamma(\rho_0, \rho_1)\big\}.$$

Proof. Since $\Gamma(\rho_0, \rho_1) \cap AC^2(0, T; \mathcal{P}_2(\mathbb{R}^d))$ is never empty, if $\inf\{A_T[\sigma(\cdot)]:\ \sigma(\cdot) \in \Gamma(\rho_0, \rho_1)\} = +\infty$, then any path in $\Gamma(\rho_0, \rho_1)$ is a minimizer.

We now assume $\inf\{A_T[\rho(\cdot)]:\ \rho(\cdot) \in \Gamma(\rho_0, \rho_1)\} < \infty$ with $\{\sigma_n(\cdot)\} \subset \Gamma(\rho_0, \rho_1)$ a minimizing sequence. Since $\|V \vee 0\|_\infty < \infty$, the paths $\sigma_n(\cdot)$s can be chosen so that $\sup_n K_T(\sigma_n(\cdot)) < \infty$. By Lemma 2.9, $\{\sigma_n(\cdot)\}$ is relatively compact in $C([0,T]; \mathcal{P}_2(\mathbb{R}^d))$. Note that K_T is lower semicontinuous by Lemma 2.9. The upper semicontinuity of V implies upper semicontinuity of

$$\int_0^T V\big(\sigma(r)\big)\,dr : C\big([0,T];\mathcal{P}_2\big(\mathbb{R}^d\big)\big) \mapsto \mathbb{R}\cup\{-\infty\}.$$

These give lower semicontinuity of A_T as a functional on $C([0,T];\mathcal{P}_2(\mathbb{R}^d))$. Choose ρ to be any limiting path of $\{\sigma_n: n=1,2,\ldots\}$, it is therefore a minimizer of A_T. The desired result follows. In particular, the $\rho\in AC^2(0,T;\mathcal{P}(\mathbb{R}^d))$ regularity follows from Lemma 2.1. □

Suppose that (1.55) holds, we now introduce another functional $J=J_T : C([0,T];\mathcal{P}_2(\mathbb{R}^d))\mapsto\mathbb{R}\cup\{+\infty\}$ by

$$J_T\big(\sigma(\cdot)\big) := \int_0^T \left[\frac{1}{2}\big\|\dot{\sigma}(s)\big\|^2_{-1,\sigma(s)} + \frac{1}{2}\nu^2 I\big(\sigma(s)\big) - V\big(\sigma(s)\big)\right] ds. \tag{3.1}$$

Note that, because of (1.55), this functional is well defined even in the absence of assumption $\|V\vee 0\|<\infty$. In the following, we will only be interested in paths with finite kinetic energy $K_T[\rho(\cdot)]<\infty$. Then, by the first part of Lemma 2.1 (see also Lemma 2.7), $S(\rho(r))<\infty$ for all $0<r\leqslant T$, regardless of the finiteness of $S(\rho_0)$. On the other hand, recall that, if

$$D(\rho_1\|\rho_0;T)<\infty, \tag{3.2}$$

by Lemma 2.10, there always exists $\hat{\sigma}(\cdot)\in\Gamma(\rho_0,\rho_1)$ minimizing K_T over all $\sigma(\cdot)\in\Gamma(\rho_0,\rho_1)$ with $K_T[\hat{\sigma}(\cdot)]<\infty$. Hence the set $\{\sigma(\cdot)\in\Gamma(\rho_0,\rho_1)\colon K_T[\sigma(\cdot)]<\infty\}$ is non-empty under this assumption.

Lemma 3.2. *Let $\rho_0,\rho_1\in\mathcal{P}_2(\mathbb{R}^d)$ satisfy* (3.2) *and $S(\rho_0)<\infty$. We assume that Condition* 1.6 *holds. Then there exists a path $\rho(\cdot)\in\Gamma(\rho_0,\rho_1)\cap AC^2(0,T;\mathcal{P}(\mathbb{R}^d))$ with*

$$\begin{aligned} A_T\big[\rho(\cdot)\big] &= \inf\big\{A_T\big[\sigma(\cdot)\big]\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1),\, K_T\big[\sigma(\cdot)\big]<\infty\big\} \\ &= \inf\big\{J_T\big(\sigma(\cdot)\big)\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1)\big\} + \nu\big(S(\rho_1)-S(\rho_0)\big) < \infty. \end{aligned} \tag{3.3}$$

Moreover, every such minimizer ρ must satisfy

$$\int_0^T \big(\big\|\dot{\rho}(s)\big\|^2_{-1,\rho(s)} + I\big(\rho(s)\big)\big)\,ds < \infty.$$

If, furthermore, $\|V\vee 0\|<\infty$, then

$$\inf\big\{A_T\big[\sigma(\cdot)\big]\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1),\, K_T\big[\sigma(\cdot)\big]<\infty\big\} = \inf\big\{A_T\big[\sigma(\cdot)\big]\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1)\big\}.$$

Proof. Let $\sigma\in\Gamma(\rho_0,\rho_1)$ with $K_T[\sigma(\cdot)]<\infty$. By Lemma 2.6,

$$\int_0^T \big(\big\|\dot{\sigma}(s)\big\|^2_{-1,\sigma(s)} + I\big(\sigma(s)\big)\big)\,ds < \infty, \tag{3.4}$$

$\sigma\in AC^2(0,T;\mathcal{P}_2(\mathbb{R}^d))$ and, with (1.55),

$$A_T\big[\sigma(\cdot)\big] = J_T\big(\sigma(\cdot)\big) + \nu\big(S(\rho_1)-S(\rho_0)\big).$$

Moreover, with (1.55), (3.4) hold when $J_T(\sigma(\cdot))<\infty$. Therefore

$$\begin{aligned} &\inf\big\{A_T\big[\sigma(\cdot)\big]\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1),\, K_T\big[\sigma(\cdot)\big]<\infty\big\} \\ &\quad= \inf\big\{J_T\big(\sigma(\cdot)\big)\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1)\big\} + \nu\big(S(\rho_1)-S(\rho_0)\big). \end{aligned}$$

To show existence of a ρ satisfying the conditions of the lemma and

$$J_T(\rho) = \inf\big\{J_T\big(\sigma(\cdot)\big)\colon \sigma(\cdot)\in\Gamma(\rho_0,\rho_1)\big\},$$

we only need to verify that J is lower semicontinuous and has compact finite level sets.

Next, we prove that J_T is lower semicontinuous. First, noting the variational nature of the $H_{-1,\rho}$ norm (see (1.18)),

$$\frac{1}{2}\int_0^T \|\dot\rho\|_{-1,\sigma}^2\,ds = \sup_{\varphi\in C_c^\infty([0,\infty)\times R^d)}\Bigg[\big\langle\varphi(T,\cdot),\rho(T)\big\rangle - \big\langle\varphi(0,\cdot),\sigma(0)\big\rangle - \int_0^T\Big\langle\partial_t\varphi+\frac{1}{2}|\nabla\varphi|^2,\rho(t)\Big\rangle\,dt\Bigg].$$

Therefore, the above is lower semicontinuous on $C([0,T];\mathcal{P}_2(\mathbb{R}^d))$ (see also Proposition 3 in [15]). Second, in view of Lemma B.1, Condition 1.6 implies that

$$\frac{\nu^2}{2}I(\rho)-V(\rho):\mathcal{P}_2\big(\mathbb{R}^d\big)\mapsto\mathbb{R}_+\cup\{+\infty\}$$

is lower semicontinuous. Therefore, the functional

$$\int_0^T\Big(\frac{\nu^2}{2}I\big(\rho(t)\big)-V\big(\rho(t)\big)\Big)\,dt$$

is also lower semicontinuous on $C([0,T];\mathcal{P}_2(\mathbb{R}^d))$. Combine the above two steps together, J_T is lower semicontinuous.

Next, we show that J has compact finite level sets. Let $\{\sigma_n(\cdot)\}\subset\Gamma(\rho_0,\rho_1)$ be such that $\sup_n J_T(\sigma_n(\cdot))<\infty$. By (1.55),

$$\sup_n\int_0^T\big[\big\|\dot\sigma_n(s)\big\|_{-1,\sigma_n(s)}^2+I\big(\sigma_n(s)\big)\big]\,ds<\infty. \tag{3.5}$$

By (2.10) in Lemma 2.6 and by Lemma 2.1,

$$\sup_n\sup_{0\leqslant s\leqslant T}S\big(\sigma_n(s)\big)<\infty.$$

Since S has compact finite level set,

$$\sigma_n(s)\in K,\quad 0\leqslant s\leqslant T,\ n=1,2,\ldots$$

for some compact set $K\subset\mathcal{P}_2(\mathbb{R}^d)$. By Theorem 8.3.1 of [2] and property of metric derivative (e.g. Theorem 1.1.2 of [2]),

$$d\big(\sigma_n(t),\sigma_n(s)\big)\leqslant\int_s^t\big\|\dot\sigma_n(r)\big\|_{-1,\sigma_n(r)}\,dr\leqslant C_T|t-s|^{\frac{1}{2}},\quad n=1,2,\ldots,\ 0\leqslant s\leqslant t\leqslant T$$

for some $C_T>0$. By Ascoli–Arzelà theorem, $\{\sigma_n(\cdot)\colon n=1,2,\ldots\}$ is relatively compact in $C([0,T];\mathcal{P}_2(\mathbb{R}^d))$. □

3.2. Minimizer of action satisfies the compressible Euler equations

Suppose that (3.2) and (1.55) hold, and that $S(\rho_0)<\infty$. Let $\rho(\cdot)\in\Gamma(\rho_0,\rho_1)\cap AC^2(0,T;\mathcal{P}_2(\mathbb{R}^d))$ be a minimizer of A_T as in (3.3) of Lemma 3.2. By the absolute continuity of $\rho(\cdot)$, and by Theorem 8.3.1 in [2], there exists a unique Borel vector field $u:(0,T)\times\mathbb{R}^d\to\mathbb{R}^d$ with $u(t,\cdot)\in L^2_{\nabla,\rho(t)}(\mathbb{R}^d)$ (see (1.13) for definition) a.e. $t\in(0,T)$ satisfying the continuity equation

$$\partial_t\rho+\operatorname{div}(\rho u)=0$$

in the sense of distribution. In this section, we prove that (Theorem 3.7) every such pair (ρ,u) is a weak solution to system (1.1) in the sense of Definition 1.9.

Let $\xi \in C_c^\infty((0,T)\times\mathbb{R}^d;\mathbb{R}^d)$ with $\xi(0,x)=\xi(T,x)=0$, $x\in\mathbb{R}^d$. First, we define $\{\eta(t,\epsilon,\cdot)\colon\ t\in[0,T],\epsilon>0\}$ a class of deformation of the identity map in direction ξ in the following sense: for each $t\in[0,T]$ fixed, let $\eta(t,\cdot,\cdot):[0,T]\times(-1,1)\times\mathbb{R}^d\mapsto\mathbb{R}^d$ be the smooth flow in ϵ given by

$$\begin{cases} \dfrac{d}{d\epsilon}\eta(t,\epsilon,x)=\xi\big(t,\eta(t,\epsilon,x)\big),\\ \eta(t,0,x)=x. \end{cases}$$

In the above, t is a parameter in the flow and ϵ is playing the role of time. By a well-known result (e.g. Theorem 2 on p. 84 of [23]), $\eta(t,\epsilon,x)\in C^\infty((0,T)\times(-1,1)\times\mathbb{R}^d)$. We will write $\eta^\epsilon(t,x):=\eta(t,\epsilon,x)$ to emphasize the role of t,x, when ϵ is viewed as a parameter; and $\eta_t^\epsilon(x):=\eta(t,\epsilon,x)$ to emphasize the role of x, when both ϵ,t are viewed as parameters. Section 4.1 of [17] contains a number of useful properties for η which we will use in the following derivations.

We now introduce a perturbation of the path $\rho(\cdot)$ along smooth direction ξ:

$$\rho^\epsilon(t):=\eta^\epsilon(t,\cdot)_\#\rho(t),\quad \text{for } t\in[0,T].$$

That is

$$\int_{\mathbb{R}^d}\varphi(y)\,\rho^\epsilon(t,dy)=\int_{\mathbb{R}^d}\varphi\big(\eta^\epsilon(t,x)\big)\rho(t,dx) \tag{3.6}$$

for every bounded measurable function φ on $\mathbb{R}^d$. $\xi(0,x)=\xi(T,x)=0$ implies that $\rho^\epsilon(0)=\rho(0)$, and $\rho^\epsilon(T)=\rho(T)$.

Lemma 3.3. $\rho^\epsilon\in AC^2(0,T;\mathcal{P}_2(\mathbb{R}^d))$.

Proof. Let $0\leqslant s<t\leqslant T$. By Kantorovich duality, there exists $\pi_{st}(dx,dy)\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$ with

$$d^2\big(\rho(s),\rho(t)\big)=\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\pi_{st}(dx,dy).$$

Let random variables $(X(s),X(t))$ be such that $\pi_{st}(dx,dy)=P(X(s)\in dx,Y(t)\in dy)$. Then $X^\epsilon(r):=\eta^\epsilon(r,X(r))$ has probability law $\rho^\epsilon(r,dx)=P(X^\epsilon(r)\in dx)$, $r=s,t$. By global Lipschitz continuity of $\eta^\epsilon(t,x)$ in (t,x), there exists $L=L_\xi>0$ so that

$$\begin{aligned} d^2\big(\rho^\epsilon(s),\rho^\epsilon(t)\big)&\leqslant E\big[\big|X^\epsilon(s)-X^\epsilon(t)\big|^2\big]\leqslant L^2\big(|s-t|^2+E\big[\big|X(s)-X(t)\big|^2\big]\big)\\ &\leqslant L^2\big[|s-t|^2+d^2\big(\rho(s),\rho(t)\big)\big]. \end{aligned}$$

The absolute continuity of ρ^ϵ now follows from the absolute continuity of ρ. $\square$

By Theorem 8.3.1 of [2], $\rho^\epsilon\in AC^2(0,T;\mathcal{P}_2(\mathbb{R}^d))$ implies the existence of a Borel field $u^\epsilon(t,x)$ with $u^\epsilon(t,\cdot)\in L^2_{\nabla,\rho^\epsilon(t)}(\mathbb{R}^d)$ (see (1.13) for definition) a.e. $t\in(0,T)$ such that

$$\partial_t\rho^\epsilon+\operatorname{div}\big(\rho^\epsilon u^\epsilon\big)=0.$$

We now characterize the time evolution of u^ϵ next. Let $\varphi\in C_c^\infty(\mathbb{R}^d)$, for ϵ sufficiently small,

$$\begin{aligned} \frac{d}{dt}\int_{\mathbb{R}^d}\varphi\,d\rho^\epsilon(t)&=\frac{d}{dt}\int_{\mathbb{R}^d}\varphi\big(\eta^\epsilon(t,x)\big)\rho(t,dx)\\ &=\int_{\mathbb{R}^d}\nabla\varphi\big(\eta^\epsilon(t,x)\big)\cdot\partial_t\eta^\epsilon(t,x)\rho(t,dx)-\int_{\mathbb{R}^d}\varphi\big(\eta^\epsilon(t,x)\big)\operatorname{div}(\rho u)\,dx\\ &=\int_{\mathbb{R}^d}\nabla\varphi\big(\eta^\epsilon(t,x)\big)\cdot\big[\partial_t\eta^\epsilon(t,x)+\big(u(t,x)\cdot\nabla\big)\eta^\epsilon(t,x)\big]\rho(t,dx)\\ &=\int_{\mathbb{R}^d}\nabla\varphi(y)\cdot\big[\partial_t\eta^\epsilon+(u\cdot\nabla)\eta^\epsilon\big]\big(t,\big(\eta^\epsilon(t,\cdot)\big)^{-1}(y)\big)\rho^\epsilon(t,dy), \end{aligned}$$

where the last equality follows from (3.6). The above sequence of displays identifies that

$$u^\epsilon(t,x) = \big[\partial_t \eta^\epsilon + (u\cdot\nabla)\eta^\epsilon\big]\big(t, \big(\eta^\epsilon(t,\cdot)\big)^{-1}(x)\big).$$

In particular, $u^0(t,x) = u(t,x)$. It follows then

$$\big\|\dot{\rho^\epsilon}(t)\big\|^2_{-1,\rho^\epsilon(t)} = \int_{\mathbb{R}^d} \big|u^\epsilon(t,x)\big|^2 \rho^\epsilon(t,dx) = \int_{\mathbb{R}^d} \big|\partial_t \eta^\epsilon(t,x) + (u\cdot\nabla)\eta^\epsilon(t,x)\big|^2 \rho(t,dx). \tag{3.7}$$

Lemma 3.4.

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} \frac{1}{2}\int_0^T \big\|\dot{\rho^\epsilon}(t)\big\|^2_{-1,\rho^\epsilon(t)}\,dt = \int_0^T\int_{\mathbb{R}^d} u(t,x)\cdot\big[\partial_t\xi(t,x) + (u\cdot\nabla)\xi(t,x)\big]\rho(t,dx)\,dt.$$

Proof. Let

$$\begin{aligned} f(t,\epsilon,x) &:= \frac{\partial}{\partial\epsilon}\frac{1}{2}\big|\partial_t\eta^\epsilon(t,x) + (u\cdot\nabla)\eta^\epsilon(t,x)\big|^2 \\ &= \big[\partial_t\eta^\epsilon(t,x) + (u\cdot\nabla)\eta^\epsilon(t,x)\big]\cdot\bigg[\frac{d}{dt}\xi\big(t,\eta^\epsilon(t,x)\big) + (u\cdot\nabla)\xi\big(t,\eta^\epsilon(t,x)\big)\bigg]. \end{aligned}$$

Let $D(\epsilon;t,x) := D_{t,x}\eta(t,\epsilon,x)$, then

$$\frac{d}{d\epsilon}D(\epsilon;t,x) = \big(0,\xi_2\big(t,\eta(t,\epsilon,x)\big)\big)^\tau\cdot D(\epsilon;t,x),$$

where $\xi_2(t,x) = D_x\xi(t,x)$. By Gronwall inequality,

$$\sup_{\epsilon\in(-1,1)}\sup_{t\in[0,T]}\big|D_{t,x}\eta^\epsilon(t,x)\big| \leqslant c\big(1+|x|\big).$$

Therefore

$$\big|f(t,\epsilon,x)\big| \leqslant C\big(1+|x|^2+\big|u(t,x)\big|^2\big)$$

with a constant C independent of (t,ϵ,x). Consequently, by (3.7) and dominated convergence theorem

$$\begin{aligned} \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\frac{1}{2}\int_0^T\big\|\dot{\rho^\epsilon}(t)\big\|^2_{-1,\rho^\epsilon(t)}\,dt &= \lim_{\delta\to0+}\frac{1}{2\delta}\int_0^T\int_{\mathbb{R}^d}\int_{\epsilon=-\delta}^{\delta} f(t,\epsilon,x)\,d\epsilon\rho(t,dx)\,dt \\ &= \int_0^T\int_{\mathbb{R}^d} f(t,0,x)\rho(t,dx)\,dt \\ &= \int_0^T\int_{\mathbb{R}^d} u(t,x)\cdot\big[\partial_t\xi(t,x) + (u\cdot\nabla)\xi(t,x)\big]\rho(t,dx)\,dt. \qquad \square \end{aligned}$$

From Lemma 3.2, we know that (2.8) holds. In view of (4.9) and (4.10) of [17],

$$\int_0^T I\big(\rho^\epsilon(t)\big)\,dt < \infty. \tag{3.8}$$

Lemma 3.5. *Assume* $\int_0^T I(\rho(t))\,dt < \infty$, *then*

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} \frac{1}{2}\int_0^T I\big(\rho^\epsilon(t)\big)\,dt = \int_0^T \int_{\mathbb{R}^d} \left(-\frac{\nabla\rho}{\rho} D\xi \frac{\nabla\rho}{\rho} + \Delta \operatorname{div}\xi + \frac{1}{2}\xi\cdot\nabla\psi\right)\rho(t,dx)\,dt.$$

Proof. By Appendix D.6 in [14],

$$\int_0^T I\big(\rho(t)\big)\,dt = 4\int_0^T \int_{\mathbb{R}^d} \left|\nabla\sqrt{\frac{d\rho(t)}{d\mu^\Psi}(x)}\right|^2 dx\,dt < \infty,$$

which implies, in particular, that $\nabla\rho(t) \in L^1_{loc}(\mathbb{R}^d)$, $\nabla\sqrt{\rho} = \frac{1}{2}\frac{\nabla\rho}{\sqrt{\rho}} \in L^2_{loc}(\mathbb{R}^d)$ almost every $t \in (0,T)$.

Because of the basic estimates in (4.24) and (4.25) and (4.19) of [17], by dominated convergence theorem,

$$\begin{aligned}
\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \int_0^T I\big(\rho^\epsilon(r)\big)\,dr &= \lim_{\delta\to 0+} \int_0^T \int_{\epsilon=-\delta}^{\delta} \int_{x\in\mathbb{R}^d} \frac{1}{2\delta} 4\frac{\partial}{\partial\epsilon}\left|\nabla\sqrt{\frac{d\rho^\epsilon(t)}{d\mu^\Psi}(x)}\right|^2 dx\,d\epsilon\,dt \\
&= \int_0^T \frac{d}{d\epsilon}\bigg|_{\epsilon=0} I\big(\rho^\epsilon(r)\big)\,dt.
\end{aligned}$$

By (4.17) in Theorem 4.1 and (4.28) in Corollary 4.1, both of [17], and by integration by parts,

$$\begin{aligned}
\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} \frac{1}{2} I\big(\rho^\epsilon(r)\big) &= \int_{\mathbb{R}^d} \left[-\left(D\xi\frac{\nabla\rho}{\sqrt{\rho}}\right)\frac{\nabla\rho}{\sqrt{\rho}} - \nabla(\operatorname{div}\xi)\nabla\rho\right]dx + \frac{1}{2}\int_{\mathbb{R}^d} \xi\cdot\nabla\psi\,d\rho \\
&= \int_{\mathbb{R}^d} \left(-\frac{\nabla\rho}{\rho} D\xi\frac{\nabla\rho}{\rho} + \Delta\operatorname{div}\xi + \frac{1}{2}\xi\cdot\nabla\psi\right)\rho(t,dx).
\end{aligned}$$

Combine the above results yields the lemma. □

By (3.8) and (1.55),

$$\int_0^T V\big(\rho^\epsilon(t)\big)\,dt < \infty.$$

Next, we show that

Lemma 3.6. *Under Condition* 1.5, *for each bounded open set* $\mathcal{O} \subset \mathbb{R}^d$,

$$\int_0^T \int_{\mathcal{O}} \big|P\big(\rho(t,x)\big)\big|\,dx\,dt \leqslant C\left(1 + \int_0^T I\big(\rho(t)\big)\,dt\right) < \infty, \tag{3.9}$$

and

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} \int_0^T V\big(\rho^\epsilon(t)\big)\,dt = \int_0^T \int_{\mathbb{R}^d} \left[\nabla\big(\phi(x) + \Phi *_x \rho(t,x)\big) + \frac{\nabla P(\rho(t,x))}{\rho(t,x)}\right]\cdot\xi(t,x)\rho(t,dx)\,dt.$$

Proof. We have

$$W\big(\rho^\epsilon(t)\big) = \int_{\mathbb{R}^d} \phi(y)\rho^\epsilon(t,dy) + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi(y_1 - y_2)\rho^\epsilon(t,dy_1)\rho^\epsilon(t,\,dy_2)$$
$$= \int_{\mathbb{R}^d} \phi\big(\eta^\epsilon(t,x)\big)\rho(t,dx) + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi\big(\eta^\epsilon(t,x_1) - \eta^\epsilon(t,x_2)\big)\rho(t,dx_1)\rho(t,dx_2).$$

This together with the fact that Φ is even gives

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_0^T W\big(\rho^\epsilon(t)\big)\,dt = \int_0^T\int_{\mathbb{R}^d}\big[\nabla\big(\phi + \Phi * \rho(t)\big)\big]\cdot\xi\,d\rho(t,dx)\,dt.$$

In the above, we applied dominated convergence theorem, which holds because that $|\frac{\partial}{\partial\epsilon}\eta^\epsilon(t,x)| \leqslant \|\xi\|_\infty < \infty$.

To consider the $\int_{\mathbb{R}^d} F(\rho^\epsilon(t))\,dt$ term, we modify the arguments of Lemma 10.4.4 in [2]. First, when ϵ is small enough (depends on ξ), we get from the area formula for push forward of probability measures (e.g. Lemma 5.5.3 of [2]) that

$$\int_{\mathbb{R}^d} F\big(\rho^\epsilon(t,y)\big)\,dy = \int_{\mathbb{R}^d} F\bigg(\frac{\rho(t,x)}{|\det D\eta^\epsilon(t,x)|}\bigg)\big|\det\eta^\epsilon(t,x)\big|\,dx = \int_{\mathbb{R}^d} G\big(\rho(t,x), \big|J(t,\epsilon,x)\big|\big)\,dx,$$

where $G(x,s) := sF(s^{-1}x)$ and $J(t,\epsilon,x) := \det D\eta^\epsilon(t,x)$ and the D means derivative in the x-variable D_x. Direct calculation (e.g. (4.5)–(4.6) of [17]) gives that

$$\partial_\epsilon J(t,\epsilon,x) = \operatorname{div}\xi\big(t,\eta(t,\epsilon,x)\big)J(t,\epsilon,x), \quad J(t,0,x) = 1;$$

and that

$$\partial_\epsilon l(t,\epsilon,x) = \operatorname{div}\xi\big(t,\eta(t,\epsilon,x)\big), \quad l(t,0,x) = 0,$$

where $l(t,\epsilon,x) := \log J(t,\epsilon,x)$. Therefore

$$\sup_{t\in[0,T],x\in\mathbb{R}^d}\big|l(t,\epsilon,x)\big| \leqslant c_1\epsilon.$$

Noting $\partial_s G(x,s) = -P(s^{-1}x)$, for $\epsilon \in (-\epsilon_0,\epsilon_0)$ for some $\epsilon_0 > 0$,

$$\frac{\partial}{\partial\epsilon}G\big(\rho(t,x), \big|J(t,\epsilon,x)\big|\big) = -P\bigg(\frac{\rho(t,x)}{J(t,\epsilon,x)}\bigg)\operatorname{div}\xi\big(t,\eta(t,\epsilon,x)\big)J(t,\epsilon,x).$$

By (1.54) and estimate (2.24), (3.9) holds and by dominated convergence theorem,

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\int_0^T\int_{\mathbb{R}^d} F\big(\rho^\epsilon(t,y)\big)\,dy\,dt = \lim_{\delta\to0+}\frac{1}{2\delta}\int_0^T\int_{\mathbb{R}^d}\int_{-\delta}^{\delta}\frac{\partial}{\partial\epsilon}G\big(\rho(t,x), J(t,\epsilon,x)\big)\,d\epsilon\,dx\,dt$$
$$= \int_0^T\int_{\mathbb{R}^d}\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}G\big(\rho(t,x), J(t,\epsilon,x)\big)\,dx\,dt$$
$$= -\int_0^T\int_{\mathbb{R}^d} P\big(\rho(t,x)\big)\operatorname{div}\xi(x)\,dx\,dt. \qquad \square$$

Combine the above results, we have

Theorem 3.7. *Assume that* (3.2) *and* (1.55) *holds, and that* $S(\rho_0) < \infty$. *Under Condition* 1.5, *any minimizer* $\sigma(\cdot) \in \Gamma(\rho_0, \rho_1)$ *of* (3.3) *in Lemma* 3.2 *satisfies*

$$\int_0^T \big(\|\dot\sigma\|_{-1,\sigma}^2 + I(\sigma)\big)\,ds < \infty \tag{3.10}$$

and is a weak solution to (1.1).

4. Wasserstein continuity of the value functions

Throughout this section, we assume that

$$\|V \vee 0\|_\infty + \|h \vee 0\|_\infty + \|g \vee 0\|_\infty < \infty. \tag{4.1}$$

Among other things, this implies that the value function f in (1.48) and the U in (1.47) are well defined. First, we give some growth estimates on f.

Lemma 4.1. *Suppose that Condition* 1.5 *holds and that Condition* 1.6 *is satisfied with* V *replaced by* h. *Then*

$$-\zeta\big(S(\rho)\big) \leqslant f(\rho) \leqslant \|h \vee 0\|_\infty + \alpha\|V \vee 0\|_\infty \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big), \tag{4.2}$$

for some sub-linear function $\zeta : \mathbb{R}_+ \mapsto \mathbb{R}_+$.

Moreover, if Condition 1.7 *is satisfied for* V *and for* h (*that is, with the* V *there replaced by* h), *then* $f : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}$ *is a finite function.*

Proof. First, we recall that

$$f(\rho_0) = \sup\left\{ \int_0^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}h(\rho) - L(\rho, \dot\rho)\big]\,ds\colon\ \rho(\cdot) \in C\big([0,\infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big),\ \rho(0) = \rho_0 \right\}.$$

The upper bound $f \leqslant \|h \vee 0\|_\infty + \alpha\|V \vee 0\|_\infty$ follows.

Next,

$$f(\rho_0) \geqslant \int_0^\infty \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\sigma(s)\big) + \alpha V\big(\sigma(s)\big)\big]\,ds, \tag{4.3}$$

where $\sigma(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ is a path satisfying

$$\sigma(0) = \rho_0 \quad \text{and} \quad \partial_t\sigma = \nu\big(\Delta\sigma + \operatorname{div}(\sigma\nabla\Psi)\big).$$

By Lemma 2.5, $\frac{\nu}{2}\int_0^t I(\sigma(r))dr \leqslant S(\rho_0)$, for every $t \geqslant 0$. By Fubini theorem,

$$\int_{t=0}^\infty \alpha^{-1}e^{-\alpha^{-1}t} I\big(\sigma(t)\big)\,dt = \alpha^{-1}\int_{t=0}^\infty \left(\alpha^{-1}e^{-\alpha^{-1}t}\int_{r=0}^t I\big(\sigma(r)\big)\,dr\right)dt \leqslant \frac{2}{\alpha\nu}S(\rho_0).$$

Under the assumptions, by Lemma 2.18, $h + \alpha V \geqslant -\hat\zeta \circ I$ for some nondecreasing sub-linear function $\hat\zeta$. We may assume without loss of generality that $\hat\zeta$ is concave. Then (4.3) yields

$$f(\rho_0) \geqslant -\int_0^\infty \alpha^{-1}e^{-\alpha^{-1}s}\hat\zeta\big(I\big(\sigma(s)\big)\big)\,ds \geqslant -\hat\zeta\left(\int_0^\infty \alpha^{-1}e^{-\alpha^{-1}r} I\big(\sigma(r)\big)\,dr\right) \geqslant -\hat\zeta\left(\frac{2}{\alpha\nu}S(\rho_0)\right),$$

where we used Jensen's inequality for the second inequality.

Now, suppose that Condition 1.7 is satisfied for V as well as for h. Then $h+\alpha V \geqslant -\hat{\zeta}\circ S$ for some nondecreasing sub-linear function $\hat{\zeta}$. By Lemma 2.7, there exists $T>0$ such that $S(\sigma(T))<\infty$, and $\int_0^T S(\sigma(s))\,ds<\infty$. In addition, Lemma 2.5 gives that $\sup_{s\geqslant T} S(\sigma(s)) \leqslant S(\sigma(T))$. Therefore, it follows from (4.3) that

$$\begin{aligned} f(\rho_0) &\geqslant -\int_0^\infty \alpha^{-1}e^{-\alpha^{-1}s}\hat{\zeta}\big(S\big(\sigma(s)\big)\big)\,ds \\ &\geqslant -\int_0^T \alpha^{-1}e^{-\alpha^{-1}s}\hat{\zeta}\big(S\big(\sigma(s)\big)\big)\,ds - \hat{\zeta}\big(S\big(\sigma(T)\big)\big)\int_T^\infty \alpha^{-1}e^{-\alpha^{-1}s}\,ds > -\infty, \end{aligned}$$

proving the finiteness of f. □

We discuss continuity of f in the following three lemmas.

Lemma 4.2. *Assume that Condition* 1.5*, and Condition* 1.6 *for h (i.e. with V replaced by h), both hold. Then for every $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ with $S(\rho_0)<\infty$, we have*

$$\liminf_{\delta\to 0+} f(P_\delta\rho_0) \geqslant f(\rho_0). \tag{4.4}$$

Assume that, in addition, Condition 1.7 *holds for V and for h. Then* (4.4) *holds for every $\rho_0\in\mathcal{P}_2(\mathbb{R}^d)$.*

Proof. Let $\epsilon>0$, by the definition of f, there exists $\hat{\rho}(\cdot)\in C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ with $\hat{\rho}(0)=\rho_0$ such that

$$f(\rho_0) \leqslant \epsilon + \int_0^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}h\big(\hat{\rho}(s)\big) - L\big(\hat{\rho}(s),\dot{\hat{\rho}}(s)\big)\big]\,ds.$$

From the finiteness of $f(\rho_0)$, $\|h\vee 0\|_\infty$ and $\|V\vee 0\|_\infty$, it follows that

$$\int_0^T T\big(\hat{\rho}(r),\dot{\hat{\rho}}(r)\big)\,dr<\infty, \quad T>0.$$

By Lemma 2.10,

$$D\big(\hat{\rho}(\tau)\big\|\rho_0;\tau\big) \leqslant \nu^{-1}\int_0^\tau T\big(\hat{\rho}(r),\dot{\hat{\rho}}(r)\big)\,dr<\infty, \quad \tau\in(0,T).$$

By Lemma 2.1, $S(\hat{\rho}(\tau))+I(\hat{\rho}(\tau))<\infty$ for almost every $\tau\in(0,T)$. For each such τ, it follows from Lemma 2.14 that $\lim_{\delta\to 0+} D(\hat{\rho}(\tau)\|P_{\nu\delta}\rho_0;\tau) = D(\hat{\rho}(\tau)\|\rho_0;\tau)<\infty$.

Let $\{\tilde{\rho}(t)\colon 0\leqslant t\leqslant\tau, \tilde{\rho}(0)=P_\delta\rho_0, \tilde{\rho}(\tau)=\hat{\rho}(\tau)\}$ be a minimizing path in (2.19) such that

$$K_\tau\big[\tilde{\rho}(\cdot)\big] = \int_0^\tau T\big(\tilde{\rho}(r),\dot{\tilde{\rho}}(r)\big)\,dr = \nu D\big(\hat{\rho}(\tau)\|P_{\nu\delta}\rho_0;\tau\big)<\infty.$$

The existence of $\tilde{\rho}$ is guaranteed by Lemma 2.10. We now extend the definition of $\tilde{\rho}$ by letting $\tilde{\rho}(t)=\hat{\rho}(t)$ for $t>\tau$. Then for each $\delta>0$, we have

$$\begin{aligned} f(P_\delta\rho_0) &\geqslant \int_0^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}h\big(\tilde{\rho}(t)\big) - L\big(\tilde{\rho}(t),\dot{\tilde{\rho}}(t)\big)\big]\,dt \\ &\geqslant \int_\tau^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}h\big(\hat{\rho}(s)\big) - L\big(\hat{\rho}(s),\dot{\hat{\rho}}(s)\big)\big]\,ds \end{aligned}$$

$$
\begin{aligned}
&+\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\tilde\rho(s)\big)+\alpha V\big(\tilde\rho(s)\big)\big]\,ds-K_\tau\big[\tilde\rho(\cdot)\big]\\
&\geqslant f(\rho_0)-\epsilon+\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\tilde\rho(s)\big)+\alpha V\big(\tilde\rho(s)\big)\big]\,ds\\
&\quad-\tau\big(\alpha^{-1}\|h\vee 0\|_\infty+\|V\vee 0\|_\infty\big)+e^{-\alpha^{-1}\tau}K_\tau\big[\hat\rho(\cdot)\big]-K_\tau\big[\tilde\rho(\cdot)\big]\\
&\geqslant f(\rho_0)-\epsilon+\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\tilde\rho(s)\big)+\alpha V\big(\tilde\rho(s)\big)\big]\,ds\\
&\quad-\tau\big(\alpha^{-1}\|h\vee 0\|_\infty+\|V\vee 0\|_\infty\big)+e^{-\alpha^{-1}\tau}\nu D\big(\hat\rho(\tau)\big\|\rho_0;\tau\big)-\nu D\big(\hat\rho(\tau)\|P_{\nu\delta}\rho_0;\tau\big).
\end{aligned}\tag{4.5}
$$

Observe that $\{\tilde\rho(\cdot)=\tilde\rho_\delta(\cdot)\colon \delta>0\}$ is relatively compact in $C([0,\tau];\mathcal{P}_2(\mathbb{R}^d))$ by Lemma 2.9. Let $\rho(\cdot)$ be a limiting path. Applying Lemma 2.7 and using the fact $\sup_{s\in[0,\tau]} d(\tilde\rho(s),\rho(s))\to 0$ as $\delta\to 0^+$, we get

$$
\limsup_{\delta\to 0^+}\int_0^\tau S\big(\tilde\rho_\delta(s)\big)\,ds\leqslant C(\tau)<\infty. \tag{4.6}
$$

Define occupation measures μ_δ,μ by

$$
\langle\varphi,\mu_\delta\rangle=\int_0^\tau \frac{e^{-\alpha^{-1}s}}{\alpha}\varphi\big(\tilde\rho_\delta(s)\big)\,ds,\qquad \langle\varphi,\mu\rangle=\int_0^\tau \frac{e^{-\alpha^{-1}s}}{\alpha}\varphi\big(\rho(s)\big)\,ds,\quad \forall\varphi\in C_b\big(\mathcal{P}_2\big(\mathbb{R}^d\big)\big).
$$

If $h+\alpha V$ satisfies Condition 1.7 (i.e. with V replaced by $h+\alpha V$), then in view of (4.6), by Lemma B.2,

$$
\begin{aligned}
&\lim_{\delta\to 0^+}\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\tilde\rho_\delta(s)\big)+\alpha V\big(\tilde\rho_\delta(s)\big)\big]\,ds\\
&\quad=\lim_{\delta\to 0^+}\int_{\mathcal{P}_2(\mathbb{R}^d)}(h+\alpha V)\,d\mu_\delta=\int_{\mathcal{P}_2(\mathbb{R}^d)}(h+\alpha V)\,d\mu\\
&\quad=\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\rho(s)\big)+\alpha V\big(\rho(s)\big)\big]\,ds.
\end{aligned}
$$

Therefore, it follows from (4.5) and Lemma 2.14 that, for a.e. $\tau\in(0,T)$,

$$
\begin{aligned}
\liminf_{\delta\to 0+} f(P_\delta\rho_0)\geqslant f(\rho_0)-\epsilon&+\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\rho(s)\big)+\alpha V\big(\rho(s)\big)\big]\,ds\\
&-\tau\big(\alpha^{-1}\|h\vee 0\|_\infty+\|V\vee 0\|_\infty\big)+\big(e^{-\alpha^{-1}\tau}-1\big)\nu D\big(\hat\rho(\tau)\|\rho_0;\tau\big).
\end{aligned}
$$

By (2.20), $\sup_{0<\tau<T}\nu D(\hat\rho(\tau)\|\rho_0;\tau)\leqslant K_T[\hat\rho(\cdot)]<\infty$. Hence the conclusion (4.4) follows by taking $\tau\to 0+$ then $\epsilon\to 0+$, and noting $\sup_{0<\tau<T}\int_0^\tau S(\rho(s))\,ds<\infty$.

Suppose that $h+\alpha V$ satisfies Condition 1.6 instead of Condition 1.7. We also assume that $S(\rho_0)<\infty$. Then by Lemma 2.5, we obtain an estimate stronger than (4.6),

$$
\limsup_{\delta\to 0^+}\int_0^\tau I\big(\tilde\rho_\delta(s)\big)\,ds\leqslant 2\nu^{-1}S(\rho_0)+2\nu^{-1}C_\nu D\big(\hat\rho(\tau)\|\rho_0;\tau\big)<\infty.
$$

In particular, $\sup_{0<\tau<T}\int_0^\tau I(\rho(s))\,ds<\infty$. Apply Lemma B.2 in Appendix B, the rest of the proof follows the same. $\square$

Lemma 4.3. *Assume that Condition* 1.5 *holds and that Condition* 1.6 *holds with V replaced by h, then f is lower semicontinuous on finite level sets of S.*

Suppose that, in addition, the stronger Condition 1.7 *holds for both V and h (i.e. with the V replaced by h), then f is lower semicontinuous on $\mathcal{P}_2(\mathbb{R}^d)$.*

Proof. Let $\rho_n \to \rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$. First, for each $\tau > 0$, we choose

$$\left\{\rho^{(n)}(t):\ 0 \leqslant t \leqslant \tau,\ \rho^{(n)}(0) = \rho_n,\ \rho^{(n)}(\tau) = P_{\nu\tau}\rho_0\right\}$$

a minimizing path in (2.19) such that

$$K_\tau\left[\rho^{(n)}(\cdot)\right] = \int_0^\tau T\left(\rho^{(n)}(r), \dot{\rho}^{(n)}(r)\right) dr = \nu D(P_{\nu\tau}\rho_0 \| \rho_n; \tau).$$

It follows from Lemma 2.14 that

$$\lim_{n\to\infty} K_\tau\left[\rho^{(n)}(\cdot)\right] = 0. \tag{4.7}$$

For each given $\epsilon > 0$, we select a path $\rho_\epsilon(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ such that $\rho_\epsilon(0) = P_{\nu\tau}\rho_0$ and

$$f(P_{\nu\tau}\rho_0) \leqslant \epsilon + \int_0^\infty e^{-\alpha^{-1}s}\left[\alpha^{-1}h(\rho_\epsilon) - L(\rho_\epsilon, \dot{\rho}_\epsilon)\right] ds.$$

We now construct a new trajectory $\sigma(\cdot) = \sigma_{n,\epsilon}(\cdot)$:

$$\sigma(t) := \rho^{(n)}(t) \quad \text{for } 0 \leqslant t \leqslant \tau, \quad \text{and} \quad \sigma(t) := \rho_\epsilon(t-\tau) \quad \text{for } t > \tau.$$

Then $\sigma \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ with $\sigma(0) = \rho_n$ and $\sigma(\tau) = P_{\nu\tau}\rho_0$. Hence

$$\begin{aligned}
f(\rho_n) &\geqslant \int_0^\infty e^{-\alpha^{-1}s}\left[\alpha^{-1}h\left(\sigma(s)\right) - L\left(\sigma(s), \dot{\sigma}(s)\right)\right] ds \\
&\geqslant \int_\tau^\infty e^{-\alpha^{-1}s}\left[\alpha^{-1}h\left(\rho_\epsilon(s-\tau)\right) - L\left(\rho_\epsilon(s-\tau), \dot{\rho}_\epsilon(s-\tau)\right)\right] ds \\
&\quad + \int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\left[h\left(\rho^{(n)}(s)\right) + \alpha V\left(\rho^{(n)}(s)\right)\right] ds - K_\tau\left[\rho^{(n)}(\cdot)\right] \\
&\geqslant e^{-\alpha^{-1}\tau}\left[f(P_{\nu\tau}\rho_0) - \epsilon\right] + \int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\left[h\left(\rho^{(n)}(s)\right) + \alpha V\left(\rho^{(n)}(s)\right)\right] ds - K_\tau\left[\rho^{(n)}(\cdot)\right]
\end{aligned}$$

for every $\epsilon > 0$. That is, taking into account of (4.7),

$$\liminf_{n\to\infty} f(\rho_n) \geqslant e^{-\alpha^{-1}\tau} f(P_{\nu\tau}\rho_0) + \liminf_{n\to\infty} \int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\left[h\left(\rho^{(n)}(s)\right) + \alpha V\left(\rho^{(n)}(s)\right)\right] ds. \tag{4.8}$$

It follows from (4.7) and Lemma 2.9 that $\{\rho^{(n)}(\cdot):\ n = 1, 2\ldots\}$ is relatively compact in $C([0,\tau]; \mathcal{P}_2(\mathbb{R}^d))$. Let $\rho^{(0)}(\cdot)$ be a limiting path. Then $\rho^{(0)}(0) = \rho_0$, and

$$\int_0^\tau T\left(\rho^{(0)}(r), \dot{\rho}^{(0)}(r)\right) dr = 0$$

by the lower semicontinuity of K_τ. Thus if we let $\bar\rho(\cdot)$ be the unique path in $C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ satisfying

$$\bar\rho(0)=\rho_0 \quad \text{and} \quad \partial_t\bar\rho=\nu\big(\Delta\bar\rho+\operatorname{div}(\bar\rho\nabla\Psi)\big),$$

then $\rho^{(0)}(s)=\bar\rho(s)$ for all $s\in[0,\tau]$.

Applying Lemma 2.7 and using the fact $\sup_{s\in[0,\tau]} d(\rho^{(n)}(s),\rho^{(0)}(s))\to 0$ as $n\to\infty$,

$$\sup_n \int_0^\tau S\big(\rho^{(n)}(s)\big)\,ds \leqslant C(\tau). \tag{4.9}$$

Suppose that $h+\alpha V$ satisfies Condition 1.7 (i.e. with V replaced by $h+\alpha V$) so that $|h+\alpha V|\leqslant \zeta^*(S)$ for some sub-linear ζ^*. As in the proof of the last lemma, introducing occupations measures and apply Lemma B.2 in Appendix B, in view of (4.9),

$$\begin{aligned}\lim_{n\to\infty}&\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\rho^{(n)}(s)\big)+\alpha V\big(\rho^{(n)}(s)\big)\big]\,ds\\ &=\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\rho^{(0)}\big)+\alpha V\big(\rho^{(0)}\big)\big]\,ds.\end{aligned} \tag{4.10}$$

Therefore (4.8) reduces to

$$\liminf_{n\to\infty} f(\rho_n)\geqslant e^{-\alpha^{-1}\tau}f(P_{\nu\tau}\rho_0)+\int_0^\tau \alpha^{-1}e^{-\alpha^{-1}s}\big[h\big(\bar\rho(s)\big)+\alpha V\big(\bar\rho(s)\big)\big]\,ds,$$

where the right hand side is a finite integral. By arbitrariness of $\tau>0$, result of the lemma follows from (4.4).

Suppose that $h+\alpha V$ satisfies Condition 1.6 instead of Condition 1.7. Then $|h+\alpha V|\leqslant\zeta^*(I)$ for some sub-linear ζ^*. If $\sup_n S(\rho_n)<\infty$, by Lemma 2.5, then

$$\sup_n\int_0^\tau I\big(\rho^{(n)}(s)\big)\,ds<\infty.$$

Apply Lemma B.2 in Appendix B, (4.10) still follows. The rest of the proof follows the same. □

Lemma 4.4. *Assume that Condition* 1.5 *holds and that Condition* 1.6 *holds for h, then f is upper semicontinuous on finite level sets of S.*

Suppose that, in addition, that stronger Condition 1.7 *holds for both V and h, then f is upper semicontinuous on $\mathcal{P}_2(\mathbb{R}^d)$.*

Proof. Let $\rho_n\to\rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$. Then there exists a sequence of $\sigma_n(\cdot)\in C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ with $\sigma_n(0)=\rho_n$ such that

$$f(\rho_n)\leqslant\frac{1}{n}+\int_0^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}h\big(\sigma_n(s)\big)-L\big(\sigma_n(s),\dot\sigma_n(s)\big)\big]\,ds.$$

We assume that $\limsup_{n\to\infty} f(\rho_n)>-\infty$ since otherwise the conclusion holds trivially. Since $L=T-V$ and $\|V\vee 0\|_\infty+\|h\vee 0\|_\infty<\infty$, selecting a subsequence if necessary,

$$\sup_n\int_0^T T\big(\sigma_n(s),\dot\sigma_n(s)\big)\,ds<\infty \quad \text{for any } T>0.$$

Therefore, $\{\sigma_n(\cdot)\colon n = 1, 2, \dots\}$ is relatively compact in $C([0,T]; \mathcal{P}_2(\mathbb{R}^d))$ by Lemma 2.9. Let $\sigma(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ be a limiting path. Then $\sigma(0) = \rho_0$.

By Lemma 2.7, for every $T > 0$, $\sup_n \int_0^T S(\sigma_n(r))\,dr < \infty$. If $V + \alpha h$ satisfies Condition 1.7, then by Lemma B.2,

$$\lim_{n\to\infty} \int_0^\infty e^{-\alpha^{-1}s}\bigl[\alpha^{-1}h\bigl(\sigma_n(s)\bigr) + V\bigl(\sigma_n(s)\bigr)\bigr]\,ds = \int_0^\infty e^{-\alpha^{-1}s}\bigl[\alpha^{-1}h\bigl(\sigma(s)\bigr) + V\bigl(\sigma(s)\bigr)\bigr]\,ds.$$

On the other hand, by lower semicontinuity,

$$\liminf_{n\to\infty} \int_0^\infty e^{-\alpha^{-1}s} T\bigl(\sigma_n(s), \dot\sigma_n(s)\bigr)\,ds \geqslant \int_0^\infty e^{-\alpha^{-1}s} T\bigl(\sigma(s), \dot\sigma(s)\bigr)\,ds.$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \int_0^\infty e^{-\alpha^{-1}s}\bigl[\alpha^{-1}h\bigl(\sigma_n(s)\bigr) - L\bigl(\sigma_n(s), \dot\sigma_n(s)\bigr)\bigr]\,ds \\
&\quad\leqslant \int_0^\infty e^{-\alpha^{-1}s}\bigl[\alpha^{-1}h\bigl(\sigma(s)\bigr) - L\bigl(\sigma(s), \dot\sigma(s)\bigr)\bigr]\,ds
\end{aligned}$$

implying

$$\limsup_{n\to\infty} f(\rho_n) \leqslant \int_0^\infty e^{-\alpha^{-1}s}\bigl[\alpha^{-1}h\bigl(\sigma(s)\bigr) - L\bigl(\sigma(s), \dot\sigma(s)\bigr)\bigr]\,ds \leqslant f(\rho_0).$$

If $V + \alpha h$ satisfies Condition 1.6 instead of Condition 1.7, and if $\sup_n S(\rho_n) < \infty$, then by Lemma 2.5,

$$\sup_n \int_0^T I\bigl(\sigma_n(r)\bigr)\,dr < \infty.$$

Noting $|V + \alpha h| \leqslant \hat\zeta(I)$ for some sub-linear function $\hat\zeta$, the rest of the arguments follow the same as before. □

We now consider the case of $U(t,\rho)$ as defined by (1.47). Similar to Lemma 4.1, we have a growth estimate for U.

Lemma 4.5. *Suppose that Conditions* 1.5, 1.8 *hold. Then*

$$-\zeta\bigl(S(\rho)\bigr) \leqslant U(t,\rho) \leqslant \|g \vee 0\|_\infty + T\|V \vee 0\|_\infty \tag{4.11}$$

for all $(t,\rho) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$, *where* $\zeta : \mathbb{R} \mapsto \mathbb{R}$ *is some sub-linear function which may depend on* $T \in (0,\infty)$. *If, additionally, Condition* 1.7 *is satisfied, then* U *is a finite function on* $[0,T] \times \mathcal{P}_2(\mathbb{R}^d)$.

The following three lemmas give continuity properties of U on $[0,\infty) \times \mathcal{P}_2(\mathbb{R}^d)$.

Lemma 4.6. *Assume that Conditions* 1.5, 1.8 *hold. Let* $t_0 > 0$ *and* $\lim_{\delta\to 0+} t_\delta \to t_0$. *Then*

$$\liminf_{\delta\to 0^+} U(t_\delta, P_\delta\rho_0) \geqslant U(t_0,\rho_0), \tag{4.12}$$

for $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *with* $S(\rho_0) < \infty$.

If in addition, we assume that Condition 1.7 *holds, then* (4.12) *holds for all* $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

Proof. Let $\epsilon > 0$. By definition of U in (1.47), there exists a $\hat{\rho}(\cdot) \in C([0,t_0];\mathcal{P}_2(\mathbb{R}^d))$ with $\hat{\rho}(0) = \rho_0$ such that

$$U(t_0,\rho_0) \leqslant \epsilon + g\big(\hat{\rho}(t_0)\big) - \int_0^{t_0} L(\hat{\rho},\dot{\hat{\rho}})\,ds.$$

Next, we extend $\hat{\rho}$ to a continuous path on $[0,\infty)$ by requiring that

$$\partial_t \hat{\rho} = -\nu \operatorname{grad} S(\hat{\rho}) \quad \text{for } t \geqslant t_0.$$

By the finiteness of $U(t_0,\rho_0)$, $\|g \vee 0\|_\infty$ and $\|V \vee 0\|_\infty$, we have

$$\int_0^T T(\hat{\rho},\dot{\hat{\rho}})\,ds < \infty.$$

Hence by Lemma 2.7, $\int_0^T S(\hat{\rho}(s))\,ds < \infty$. Also, Lemma 2.1 allows us to conclude that $S(\hat{\rho}(\tau)) < \infty$ for every $\tau \in (0,T)$ and $I(\hat{\rho}(\tau)) < \infty$ for almost every $\tau \in (0,T)$. In the case when Condition 1.7 holds, the above estimates implies $\int_0^T |V(\hat{\rho}(s))|\,ds < \infty$. In the case that Condition 1.7 is not assumed but $S(\rho_0) < \infty$, we have $\int_0^T I(\hat{\rho}(s))\,ds < \infty$ from Lemma 2.5; by Lemma 2.18, $\int_0^T |V(\hat{\rho}(s)|\,ds < \infty$ holds as well.

For each $0 < \tau < \frac{t_0}{2}$ with $I(\hat{\rho}(\tau)) < \infty$, we introduce $\{\tilde{\rho}(t)\colon 0 \leqslant t \leqslant \tau\}$ a minimizing path of K_τ on $\Gamma(P_\delta\rho_0, \hat{\rho}(\tau))$, and then extend it to all time by setting $\tilde{\rho}(t) := \hat{\rho}(t)$ for $t > \tau$. It follows then

$$\begin{aligned}
U(t_\delta, P_\delta\rho_0) &\geqslant g\big(\tilde{\rho}(t_\delta)\big) - \int_0^{t_\delta} L(\tilde{\rho},\dot{\tilde{\rho}})\,ds \\
&= g\big(\hat{\rho}(t_\delta)\big) - \int_0^{t_\delta} L(\hat{\rho},\dot{\hat{\rho}})\,ds + \int_0^{\tau} L(\hat{\rho},\dot{\hat{\rho}})\,ds - \int_0^{\tau} L(\tilde{\rho},\dot{\tilde{\rho}})\,ds \\
&\geqslant g\big(\hat{\rho}(t_\delta)\big) - \int_0^{t_\delta} L(\hat{\rho},\dot{\hat{\rho}})\,ds + \int_0^{\tau} V(\tilde{\rho})\,ds - \int_0^{\tau} V(\hat{\rho})\,ds \\
&\quad + \nu D\big(\hat{\rho}(\tau)\|\rho_0;\tau\big) - \nu D\big(\hat{\rho}(\tau)\|P_\delta\rho_0;\tau\big).
\end{aligned}$$

As in the proof of Lemma 4.3, $\{\tilde{\rho}(\cdot) = \tilde{\rho}_\delta(\cdot)\colon \delta > 0\}$ is relatively compact in $C([0,\tau];\mathcal{P}_2(\mathbb{R}^d))$. Let $\rho(\cdot) \in C([0,\tau];\mathcal{P}_2(\mathbb{R}^d))$ be a limiting path. By Lemmas 2.7 and 2.14,

$$\limsup_{\delta\to 0^+} \int_0^{\tau} S\big(\tilde{\rho}_\delta(s)\big)\,ds < \infty.$$

If Condition 1.7 holds, then combined with Lemma 2.14,

$$\begin{aligned}
\liminf_{\delta\to 0+} U(t_\delta, P_\delta\rho_0) &\geqslant g\big(\hat{\rho}(t_0)\big) - \int_0^{t_0} L(\hat{\rho},\dot{\hat{\rho}})\,ds + \int_0^{\tau} V\big(\rho(s)\big)\,ds - \int_0^{\tau} V\big(\hat{\rho}(s)\big)\,ds \\
&\geqslant g\big(\hat{\rho}(t_0)\big) - \int_0^{t_0} L(\hat{\rho},\dot{\hat{\rho}})\,ds + \int_0^{\tau} V\big(\rho(s)\big)\,ds - \|V \vee 0\|_\infty \tau.
\end{aligned} \tag{4.13}$$

If Condition 1.7 is not assumed, but $S(\rho_0) < \infty$, then

$$\limsup_{\delta\to 0^+} \int_0^{\tau} I\big(\tilde{\rho}_\delta(s)\big)\,ds < \infty$$

by Lemma 2.5. Hence (4.13) also follows by Lemma 2.18 and Lemma B.2 in Appendix B.

We have

$$K_\tau\big[\rho(\cdot)\big] \leqslant \lim_{\delta\to0+} K_\tau\big[\tilde\rho_\delta(\cdot)\big] = \lim_{\delta\to0+} \nu D\big(\hat\rho(\tau)\|P_\delta\rho_0;\tau\big) = \nu D\big(\hat\rho(\tau)\|\rho_0;\tau\big) \leqslant K_T\big[\hat\rho(\cdot)\big] < \infty.$$

Combined with Lemma 2.7,

$$\sup_{0<\tau<\frac{t_0}{2}} \int_0^\tau S\big(\rho_\tau(r)\big)\,dr < \infty.$$

Therefore, if Condition 1.7 holds, then $\lim_{\tau\to0+}\int_0^\tau V(\rho_\tau(r))\,dr = 0$; if Condition 1.7 does not hold but $S(\rho_0) < \infty$, then by Lemma 2.5,

$$\sup_{0<\tau<\frac{t_0}{2}} \int_0^\tau I\big(\rho_\tau(r)\big)\,dr < \infty.$$

Consequently $\lim_{\tau\to0+}\int_0^\tau V(\rho_\tau(r))\,dr = 0$ follows from Lemmas 2.18 and B.2.

Letting $\tau \to 0+$ in (4.13),

$$\liminf_{\delta\to0+} U(t_\delta, P_\delta\rho_0) \geqslant g\big(\hat\rho(t_0)\big) - \int_0^{t_0} L(\hat\rho,\dot{\hat\rho})\,ds \geqslant U(t_0,\rho_0) - \epsilon.$$

Conclusion of the lemma follows by taking $\epsilon \to 0+$. □

Lemma 4.7. *Assume that Conditions* 1.5, 1.8 *hold. Then $U(t,\rho)$ is lower semicontinuous on* $[0,\infty)\times K_C$ *for every* $C\in\mathbb{R}$, *where*

$$K_C := \big\{\rho\in\mathcal{P}_2\big(\mathbb{R}^d\big):\ S(\rho)\leqslant C\big\}.$$

Suppose additionally, that Condition 1.7 *holds, then $U(t,\rho)$ is lower semicontinuous on* $[0,\infty)\times\mathcal{P}_2(\mathbb{R}^d)$.

Proof. Let $\rho_n\to\rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$ in $\mathcal{P}_2(\mathbb{R}^d)$ and let $t_n\to t_0$. We will prove that

$$\liminf_{n\to\infty} U(t_n,\rho_n) \geqslant U(t_0,\rho_0).$$

We first consider the case of $t_0=0$. For each n, let $\sigma_n(\cdot)$ be a path in $C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ satisfying

$$\sigma_n(0)=\rho_n \quad\text{and}\quad \partial_t\sigma_n = -\nu\,\mathrm{grad}\,S(\sigma_n).$$

Then we have

$$\liminf_{n\to\infty} U(t_n,\rho_n) \geqslant \liminf_{n\to\infty}\left[g\big(\sigma_n(t_n)\big) + \int_0^{t_n} V\big(\sigma_n(s)\big)\,ds\right]. \tag{4.14}$$

By Lemma 2.7,

$$\sup_n \int_0^T S\big(\sigma_n(s)\big)\,ds < \infty.$$

If Condition 1.7 holds, then $\lim_{n\to\infty}\int_0^{t_n}|V(\sigma_n)|\,ds = 0$. If Condition 1.7 does not hold but $\sup_n S(\rho_n)<\infty$, then by Lemma 2.5,

$$\sup_n \int_0^T I\big(\sigma_n(s)\big)\,ds < \infty.$$

Hence $\lim_{n\to\infty}\int_0^{t_n}|V(\sigma_n)|\,ds = 0$ by Lemma 2.18.

Also, elementary properties of heat equation show that $\{\sigma_n(\cdot)\colon n = 1, 2, \ldots\}$ is a convergent sequence in $C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$, implying in particular $\sigma_n(t_n) \to \rho_0$ in $\mathcal{P}_2(\mathbb{R}^d)$. Apply these observations to (4.14), $\liminf_{n\to\infty} U(t_n, \rho_n) \geqslant g(\rho_0) = U(0, \rho_0)$ as desired.

It remains to consider the case $t_0 > 0$. Let $0 < \tau < \frac{t_0}{2}$, and assume without loss of generality that $|t_n - t_0| < \tau$ for all n. We select $\rho^{(n)}(\cdot)$ as in the beginning paragraph in the proof of Lemma 4.3, and select $\rho_\epsilon(\cdot) \in C([0, t_0 - \tau]; \mathcal{P}_2(\mathbb{R}^d))$, $\epsilon > 0$ such that $\rho_\epsilon(0) = P_{\nu\tau}\rho_0$ and

$$U(t_0 - \tau, P_{\nu\tau}\rho_0) \leqslant \epsilon + g\big(\rho_\epsilon(t_0 - \tau)\big) - \int_0^{t_0-\tau} L(\rho_\epsilon, \dot\rho_\epsilon)\, ds.$$

Next, we extend ρ_ϵ to be a continuous path on $[0, \infty)$ by requiring that

$$\partial_t \rho_\epsilon = -\nu \operatorname{grad} S(\rho_\epsilon) \quad \text{for } t \geqslant t_0 - \tau.$$

By finiteness of $U(t_0 - \tau, P_{\nu\tau}\rho_0)$ and $\|g \vee 0\|_\infty$ and $\|V \vee 0\|_\infty$, we have $K_{t_0-\tau}[\rho_\epsilon(\cdot)] < \infty$ and hence $K_T[\rho_\epsilon(\cdot)] < \infty$ for every $T > 0$. By Lemma 2.7, $\int_0^T S(\rho_\epsilon(s))\, ds < \infty$. If $S(\rho_0) < \infty$, we even have $\int_0^T I(\rho_\epsilon(s))\, ds < \infty$ by Lemma 2.5. Consequently, with the assumptions of this lemma, $\int_0^T |L(\rho_\epsilon(s), \dot\rho_\epsilon(s))|\, ds < \infty$.

We construct a new path $\sigma(\cdot) = \sigma_{n,\epsilon}(\cdot)$ by concatenating $\rho^{(n)}(\cdot)$ and $\rho_\epsilon(\cdot)$ together:

$$\sigma(t) := \rho^{(n)}(t) \quad \text{for } 0 \leqslant t \leqslant \tau, \quad \text{and} \quad \sigma(t) := \rho_\epsilon(t - \tau) \quad \text{for } t \geqslant \tau.$$

Then

$$\begin{aligned} U(t_n, \rho_n) &\geqslant g\big(\sigma(t_n)\big) - \int_0^{t_n} L(\sigma, \dot\sigma)\, ds \\ &= g\big(\rho_\epsilon(t_n - \tau)\big) - \int_0^{t_n-\tau} L(\rho_\epsilon, \dot\rho_\epsilon)\, ds + \int_0^\tau V\big(\rho^{(n)}(s)\big)\, ds - \nu D(P_{\nu\tau}\rho_0 \| \rho_n; \tau). \end{aligned} \tag{4.15}$$

As in the proof of Lemma 4.3, $\{\rho^{(n)}(\cdot)\colon n = 1, 2 \ldots\}$ is relatively compact in $C([0, \tau]; \mathcal{P}_2(\mathbb{R}^d))$, and by Lemma 2.7,

$$\sup_n \int_0^\tau S\big(\rho^{(n)}(s)\big)\, ds \leqslant C(\tau).$$

Let $\rho^{(0)}(\cdot)$ be a limiting path, then $\rho^{(0)}(s) = \bar\rho(s)$, $s \in [0, \tau]$, where $\bar\rho(\cdot)$ is the unique path in $C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ satisfying

$$\bar\rho(0) = \rho_0 \quad \text{and} \quad \partial_t \bar\rho = \nu\big(\Delta\bar\rho + \operatorname{div}(\bar\rho \nabla\Psi)\big).$$

Suppose that Condition 1.7 holds, then from Lemma 2.14 and Lemma B.2, (4.15) gives

$$\begin{aligned} \liminf_{n\to\infty} U(t_n, \rho_n) &\geqslant g\big(\rho_\epsilon(t_0 - \tau)\big) - \int_0^{t_0-\tau} L(\rho_\epsilon, \dot\rho_\epsilon)\, ds + \int_0^\tau V\big(\bar\rho(s)\big)\, ds \\ &\geqslant U(t_0 - \tau, P_{\nu\tau}\rho_0) - \epsilon + \int_0^\tau V\big(\bar\rho(s)\big)\, ds. \end{aligned}$$

This gives

$$\liminf_{n\to\infty} U(t_n, \rho_n) \geqslant \liminf_{\tau\to 0+} U(t_0 - \tau, P_{\nu\tau}\rho_0) \geqslant U(t_0, \rho_0),$$

where we used (4.12) in the last inequality.

If Condition 1.7 does not hold but $\sup_n S(\rho_n) < \infty$, then by Lemma 2.5,

$$\sup_n \int_0^T I\big(\rho^{(n)}(s)\big)\,ds < \infty.$$

Then the above arguments still go through because of Lemma 2.18. □

Lemma 4.8. *Assume that Conditions* 1.5, 1.8 *hold. Then U is upper semicontinuous on* $[0, \infty) \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant C\}$ *for every* $C \in \mathbb{R}$.

Suppose additionally, that Condition 1.7 *holds, then U is upper semicontinuous on* $[0, \infty) \times \mathcal{P}_2(\mathbb{R}^d)$.

Proof. The proof follows similarly to the proof of Lemma 4.4.

Let $(t_0, \rho_0) \in [0, \infty) \times \mathcal{P}_2(\mathbb{R}^d)$ and $t_n \to t_0$, $\rho_n \to \rho_0$. There exists a sequence of $\sigma_n(\cdot) \in C([0, t_n]; \mathcal{P}_2(\mathbb{R}^d))$ with $\sigma_n(0) = \rho_n$ such that

$$U(t_n, \rho_n) \leqslant \frac{1}{n} + g\big(\sigma_n(t_n)\big) - \int_0^{t_n} L\big(\sigma_n(s), \dot{\sigma}_n(s)\big)\,ds.$$

For each n, we extend σ_n to a continuous path on $[0, \infty)$ by

$$\partial_t \sigma_n = -\nu \operatorname{grad} S(\sigma_n) \quad \text{for } t \geqslant t_n.$$

Without loss of generality, we assume that $\limsup_{n\to\infty} U(t_n, \rho_n) > -\infty$. Then by the upper boundedness of g and V, selecting a subsequence if necessary,

$$\sup_n \int_0^T T\big(\sigma_n(s), \dot{\sigma}_n(s)\big)\,ds < \infty.$$

Therefore, $\{\sigma_n(\cdot)\colon\ n = 1, 2, \ldots\}$ is relatively compact in $C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ by Lemma 2.9. Let $\sigma(\cdot) \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ be a limiting trajectory. Then $\sigma(0) = \rho_0$ and $\lim_{n\to\infty} \sigma_n(t_n) = \sigma(t_0)$. By lower semicontinuity, $\int_0^T T(\sigma, \dot{\sigma})\,ds < \infty$. By Lemma 2.7,

$$\sup_n \int_0^T S\big(\sigma_n(r)\big)\,dr + \int_0^T S\big(\sigma(r)\big)\,dr < \infty. \tag{4.16}$$

In addition, if $\sup_n S(\rho_n) < \infty$ holds, then by Lemma 2.5, we even have

$$\sup_n \int_0^T I\big(\sigma_n(r)\big)\,dr + \int_0^T I\big(\sigma(r)\big)\,dr < \infty. \tag{4.17}$$

If $t_0 = 0$, then

$$\limsup_{n\to\infty} U(t_n, \rho_n) \leqslant g(\rho_0) + \lim_{n\to\infty} t_n \|V \vee 0\|_\infty = g(\rho_0) = U(0, \rho_0).$$

Suppose that $t_0 > 0$, it follows from (4.16) or (4.17), the assumptions relating V and S or V and I, and the fact K_T is lower semicontinuous that

$$\liminf_{n\to\infty} \int_0^{t_0} L(\sigma_n, \dot{\sigma}_n)\,ds \geqslant \int_0^{t_0} L(\sigma, \dot{\sigma})\,ds.$$

Therefore

$$\limsup_{n\to\infty} U(t_n,\rho_n) \leqslant \limsup_{n\to\infty}\left\{ g\big(\sigma_n(t_n)\big) - \int_0^{t_n} L(\sigma_n,\dot\sigma_n)\,ds \right\}$$

$$\leqslant g\big(\sigma(t_0)\big) - \int_0^{t_0} L\big(\sigma(s),\dot\sigma(s)\big)\,ds - \liminf_{n\to\infty}\int_{t_0}^{t_n} L(\sigma_n,\dot\sigma_n)\,ds. \tag{4.18}$$

If $t_n \geqslant t_0$, then

$$\liminf_{n\to\infty}\int_{t_0}^{t_n} L(\sigma_n,\dot\sigma_n)\,ds \geqslant -\lim_{n\to\infty}|t_n-t_0|\|V\vee 0\|_\infty = 0.$$

On the other hand, if $t_n < t_0$, since $T(\sigma_n(s),\dot\sigma_n(s)) = 0$ for $s \geqslant t_n$,

$$\lim_{n\to\infty}\int_{t_0}^{t_n} L(\sigma_n,\dot\sigma_n)\,ds = \lim_{n\to\infty}\int_{t_n}^{t_0} V(\sigma_n)\,ds = 0,$$

by our assumptions and (4.16)–(4.17).

Consequently,

$$\limsup_{n\to\infty} U(t_n,\rho_n) \leqslant g\big(\sigma(t_0)\big) - \int_0^{t_0} L\big(\sigma(s),\dot\sigma(s)\big)\,ds \leqslant U(t_0,\rho_0). \qquad \square$$

5. Hamilton–Jacobi equations

Let H be defined according to (1.44).

Lemma 5.1. *Let $f_0 \in D_0$ and $f_1 \in D_1$. Under Condition* 1.5*,*

(1) $Hf_0 : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}\cup\{-\infty\}$ *is upper semicontinuous and bounded from above; for each $C \in \mathbb{R}$,*

$$\big\{\rho \in \mathcal{P}_2\big(\mathbb{R}^d\big):\ Hf_0(\rho) \geqslant C\big\}$$

is compact in $\mathcal{P}_2(\mathbb{R}^d)$;

(2) $Hf_1 : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}\cup\{+\infty\}$ *is lower semicontinuous and bounded from below*; *for each $C \in \mathbb{R}$,*

$$\big\{\gamma \in \mathcal{P}_2\big(\mathbb{R}^d\big):\ Hf_1(\gamma) \leqslant C\big\}$$

is compact in $\mathcal{P}_2(\mathbb{R}^d)$.

Proof. We only show the case of Hf_0, the other case is similar.

First,

$$\begin{aligned} Hf_0(\rho) &= \frac{\theta^2}{2}d^2(\rho,\gamma) + (\epsilon-\nu)\theta\Big\langle \operatorname{grad} S(\rho), \operatorname*{grad}_\rho \frac{1}{2}d^2(\rho,\gamma)\Big\rangle_{-1,\rho} - \frac{\epsilon(2\nu-\epsilon)}{2}I(\rho) + V(\rho) \\ &:= G(\rho) - \frac{\epsilon(2\nu-\epsilon)}{2}I(\rho). \end{aligned}$$

We claim that there exists a sub-linear function $\zeta^* : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that

$$\big|G(\rho)\big| \leqslant \zeta^*\big(I(\rho)\big). \tag{5.1}$$

Then the conclusion follows from Lemma B.1, if G is continuous on level sets of I.

By assumption (1.52) and $I(\rho) \geqslant \int_{\mathbb{R}^d} \psi \, d\rho$,

$$d^2(\rho, \gamma) \leqslant 2\left[\int_{\mathbb{R}^d} |x|^2 d\gamma + \int_{\mathbb{R}^d} |y|^2 \, d\rho\right] \leqslant \zeta^*\big(I(\rho)\big)$$

for some sub-linear function ζ^*. Therefore, (5.1) follows from the above estimate, (1.55) and (1.41).

To see that G is continuous on finite level sets of I, we let $\rho_n \to \rho_0$ be such that $\sup_n I(\rho_n) < \infty$. By Lemma 2.18, $\lim_{n\to\infty} V(\rho_n) = V(\rho_0)$. On the other hand, by Lemma D.48 in [14], we also have

$$\begin{aligned}
&\lim_{n\to\infty}\left\langle \operatorname{grad} S(\rho_n), \operatorname*{grad}_\rho \frac{1}{2} d^2(\rho_n, \gamma)\right\rangle_{-1,\rho_n} \\
&\quad = \lim_{n\to\infty} \int_{\mathbb{R}^d} \nabla p_{\rho_n,\gamma} \cdot \left(\frac{\nabla \rho_n}{\rho_n} + \nabla \Psi\right) d\rho_n \\
&\quad = \int_{\mathbb{R}^d} \nabla p_{\rho_0,\gamma} \cdot \left(\frac{\nabla \rho_0}{\rho_0} + \nabla \Psi\right) d\rho_0 = \left\langle \operatorname{grad} S(\rho_0), \operatorname*{grad}_\rho \frac{1}{2} d^2(\rho_0, \gamma)\right\rangle_{-1,\rho_0}.
\end{aligned}$$

Combining the above estimates, $\lim_{n\to\infty} G(\rho_n) = G(\rho_0)$. $\square$

The following is Theorem D.50 on p. 397 of Feng and Kurtz [14], which generalizes a result of Cordero–Erausquin, Gangbo and Houdré [5].

Lemma 5.2. *Let $\rho, \gamma \in \mathcal{P}_2(\mathbb{R}^d)$ with $I(\rho) < +\infty$, then*

$$\begin{aligned}
S(\gamma) &\geqslant S(\rho) - \int_{\mathbb{R}^d} \nabla p_{\rho,\gamma} \cdot \left(\frac{\nabla \rho}{\rho} + \nabla \Psi\right) d\rho + \frac{\lambda_\Psi}{2} d^2(\rho, \gamma) \\
&= S(\rho) - \left\langle \operatorname{grad} S(\rho), \operatorname*{grad}_\rho \frac{1}{2} d^2(\rho, \gamma)\right\rangle_{-1,\rho} + \frac{\lambda_\Psi}{2} d^2(\rho, \gamma).
\end{aligned}$$

In particular, if $I(\rho) + I(\gamma) < +\infty$,

$$\left\langle \operatorname{grad} S(\rho), \operatorname{grad}_\rho \frac{1}{2} d^2(\rho, \gamma)\right\rangle_{-1,\rho} + \left\langle \operatorname{grad} S(\gamma), \operatorname*{grad}_\gamma \frac{1}{2} d^2(\rho, \gamma)\right\rangle_{-1,\gamma} \geqslant \lambda_\Psi \, d^2(\rho, \gamma).$$

From the above result, we can derive the so-called HWI inequality. See Corollary D.52 in [14].

Lemma 5.3. *For every $\rho, \gamma \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$S(\rho) \leqslant S(\gamma) + d(\rho, \gamma)\sqrt{I(\rho)} - \frac{\lambda_\Psi}{2} \, d^2(\rho, \gamma)$$

Remark 5.4. Note that $f - \alpha H f = h$ is equivalent to $f - \alpha(Hf + \alpha^{-1} h) = 0$, and that if h satisfies Condition 1.6 with the V replaced by h, then under Condition 1.5, by Lemma 2.18, $V^h := V + \alpha^{-1} h$ also satisfy the same condition. Therefore, we may assume $h = 0$ with no loss of generality in the following existence and uniqueness proofs for resolvent equation (1.45).

5.1. Uniqueness of viscosity solution to the Hamilton–Jacobi equations

5.1.1. The resolvent equation

Lemma 5.5 *(Comparison principle). Assume that Condition* 1.5 *holds and that Condition* 1.6 *is satisfied for h, and that $\|h \vee 0\|_\infty < \infty$. Let $\overline{f}, \underline{f} : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{\pm\infty\}$ be respective viscosity sub- and super-solution to* (1.45)*. Suppose further that there exist functions $\zeta_1^*, \zeta_2^* : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with sub-linear growth at infinity such that*

$$\overline{f}(\rho) \leqslant \zeta_1^*\big(S(\rho)\big) \quad \textit{and} \quad -\zeta_2^*\big(S(\rho)\big) \leqslant \underline{f}(\rho) \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big).$$

Then,

$$\overline{f}(\rho) \leqslant \underline{f}(\rho) \quad \textit{for every } \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big) \textit{ satisfying } S(\rho) < \infty.$$

If in addition $\overline{f}, \underline{f} \in C(\mathcal{P}_2(\mathbb{R}^d))$, then the above inequality holds for all ρ in $\mathcal{P}_2(\mathbb{R}^d)$. Furthermore, there is at most one viscosity solution $f \in C(\mathcal{P}_2(\mathbb{R}^d))$ to (1.45) *satisfying*

$$\big|f(\rho)\big| \leqslant \zeta^*\big(S(\rho)\big) \quad \forall \rho \in \mathcal{P}_2\big(\mathbb{R}^d\big)$$

for some function $\zeta^* : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with sub-linear growth at infinity.

Proof. As commented in Remark 5.4, we only need to prove the case $h = 0$.

Let $\lambda > 1$ and

$$G(\rho, \gamma) := \lambda \overline{f}(\rho) - \underline{f}(\gamma) - \frac{\theta}{2} d^2(\rho, \gamma) - \epsilon S(\rho) - \epsilon S(\gamma), \tag{5.2}$$

where $\theta, \epsilon > 0$. By assumption on continuity of $\overline{f}$, $\underline{f}$ on finite level sets of S, and by assumption on growth estimate of $\overline{f}$ and $\underline{f}$, the function $G : \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R} \cup \{+\infty\}$ has compact finite level sets in $\mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$, and is continuous on such level sets. Therefore there exists $(\rho_0, \gamma_0) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ such that

$$G(\rho_0, \gamma_0) = \sup_{(\rho,\gamma) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)} G(\rho, \gamma).$$

Note that $\{(\rho_0, \gamma_0)\colon \theta > 0\}$ is a relatively compact set in $\mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ when $\epsilon > 0$ is fixed. An adaptation of Proposition 3.7 in Crandall, Ishii and Lions [8] (e.g. Lemma 9.2 in [14]) gives

$$\limsup_{\theta \to +\infty} \theta\, d^2(\rho_0, \gamma_0) = 0.$$

In particular, by working with a subsequence we can assume that $\rho_0, \gamma_0 \to \rho_*$ in $\mathcal{P}_2(\mathbb{R}^d)$ as θ tends to $+\infty$.

Let

$$f_0(\rho) := \frac{\theta}{2} d^2(\rho, \gamma_0) + \epsilon S(\rho) \quad \text{and} \quad f_1(\gamma) := -\frac{\theta}{2} d^2(\rho_0, \gamma) - \epsilon S(\gamma).$$

Then ρ_0 is a maximum point of $\overline{f} - \lambda^{-1} f_0$. Hence by the sub-solution property,

$$\alpha^{-1}\big[\overline{f}(\rho_0) - h(\rho_0)\big] \leqslant H\big(\lambda^{-1} f_0\big)(\rho_0).$$

Similarly,

$$\alpha^{-1}\big[\underline{f}(\gamma_0) - h(\gamma_0)\big] \geqslant H f_1(\gamma_0).$$

It follows then for each $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned}
\lambda \overline{f}(\rho) - \underline{f}(\rho) - 2\epsilon S(\rho) &\leqslant G(\rho, \rho) \leqslant G(\rho_0, \gamma_0) \\
&\leqslant \lambda \overline{f}(\rho_0) - \underline{f}(\gamma_0) \\
&\leqslant \lambda h(\rho_0) - h(\gamma_0) + \alpha\big[\lambda H\big(\lambda^{-1} f_0\big)(\rho_0) - H f_1(\gamma_0)\big].
\end{aligned} \tag{5.3}$$

If $I(\rho_0) + I(\gamma_0) = \infty$ for some $\theta > 0$, then it follows from the definition of H in (1.44) that the right hand side above is $-\infty$, and we conclude the lemma. When $I(\rho_0) + I(\gamma_0) < \infty$ for all θ, we have

$$\begin{aligned}
&\lambda H\big(\lambda^{-1} f_0\big)(\rho_0) - H f_1(\gamma_0) \\
&\quad = -\nu\big[\langle \operatorname{grad} S(\rho_0), \operatorname{grad} f_0(\rho_0)\rangle_{-1,\rho_0} - \langle \operatorname{grad} S(\gamma_0), \operatorname{grad} f_1(\gamma_0)\rangle_{-1,\gamma_0}\big] \\
&\qquad + \frac{1}{2\lambda}\big\| \operatorname{grad} f_0(\rho_0)\big\|^2_{-1,\rho_0} - \frac{1}{2}\big\| \operatorname{grad} f_1(\gamma_0)\big\|^2_{-1,\gamma_0} + \lambda V(\rho_0) - V(\gamma_0)
\end{aligned}$$

$$
\begin{aligned}
&= -\nu\theta\Big[\Big\langle \operatorname{grad} S(\rho_0), \operatorname*{grad}_{\rho_0} \frac{1}{2} d^2(\rho_0,\gamma_0)\Big\rangle_{-1,\rho_0} + \Big\langle \operatorname{grad} S(\gamma_0), \operatorname*{grad}_{\gamma_0} \frac{1}{2} d^2(\rho_0,\gamma_0)\Big\rangle_{-1,\gamma_0}\Big] \\
&\quad - \epsilon\nu\big[I(\rho_0)+I(\gamma_0)\big] \\
&\quad + \frac{\theta^2}{2\lambda}\Big\|\operatorname*{grad}_{\rho_0}\frac{1}{2}d^2(\rho_0,\gamma_0)\Big\|^2_{-1,\rho_0} + \frac{\epsilon\theta}{\lambda}\Big\langle \operatorname*{grad}_{\rho_0}\frac{1}{2}d^2(\rho_0,\gamma_0), \operatorname{grad} S(\rho_0)\Big\rangle_{-1,\rho_0} + \frac{\epsilon^2}{2\lambda} I(\rho_0) \\
&\quad - \frac{\theta^2}{2}\Big\|\operatorname*{grad}_{\gamma_0}\frac{1}{2}d^2(\rho_0,\gamma_0)\Big\|^2_{-1,\gamma_0} - \epsilon\theta\Big\langle \operatorname*{grad}_{\gamma_0}\frac{1}{2}d^2(\rho_0,\gamma_0), \operatorname{grad} S(\gamma_0)\Big\rangle_{-1,\gamma_0} - \frac{\epsilon^2}{2} I(\gamma_0) \\
&\quad + \lambda V(\rho_0) - V(\gamma_0) \\
&\leqslant -\lambda_\Psi \nu\theta\, d^2(\rho_0,\gamma_0) - \epsilon\nu\big[I(\rho_0)+I(\gamma_0)\big] - \frac{\lambda-1}{\lambda}\frac{\theta^2}{2} d^2(\rho_0,\gamma_0) \\
&\quad + \theta d(\rho_0,\gamma_0)\epsilon\big[\sqrt{I(\rho_0)}+\sqrt{I(\gamma_0)}\big] + \frac{\epsilon^2}{2} I(\rho_0) + \lambda V(\rho_0) - V(\gamma_0) \\
&\leqslant -\lambda_\Psi \nu\theta d^2(\rho_0,\gamma_0) - \epsilon\Big[\nu - \frac{3\lambda-1}{\lambda-1}\frac{\epsilon}{2}\Big]\big[I(\rho_0)+I(\gamma_0)\big] + \lambda V(\rho_0) - V(\gamma_0).
\end{aligned}
\tag{5.4}
$$

In the above, the first inequality follows from Lemma 5.2 and (1.41).

Combined with (5.3), therefore, for each $\lambda > 1$ fixed, when $\epsilon \leqslant \frac{\nu(\lambda-1)}{3\lambda-1}$, for every $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ such that $S(\rho) < \infty$,

$$
\begin{aligned}
\lambda \overline{f}(\rho) - \underline{f}(\rho) \leqslant\ & 2\epsilon S(\rho) - \alpha\lambda_\Psi \nu\theta\, d^2(\rho_0,\gamma_0) - \alpha\epsilon\frac{\nu}{2}\big[I(\rho_0)+I(\gamma_0)\big] \\
& + \alpha(\lambda-1)\|V\vee 0\|_\infty + \alpha\big[V(\rho_0)-V(\gamma_0)\big] \\
& + (\lambda-1)\|h\vee 0\|_\infty + h(\rho_0) - h(\gamma_0).
\end{aligned}
\tag{5.5}
$$

By Lemma 2.18, I is the only dominating term on the right hand side of the above inequality.

If $\limsup_{\theta\to\infty} I(\rho_0)+I(\gamma_0) < \infty$, then by Lemma 2.18 and Condition 1.6 on h, we obtain from (5.5) by sending θ to $+\infty$ that

$$
\lambda \overline{f}(\rho) - \underline{f}(\rho) \leqslant 2\epsilon S(\rho) + \alpha(\lambda-1)\|V\vee 0\|_\infty + (\lambda-1)\|h\vee 0\|_\infty. \tag{5.6}
$$

If $\limsup_{\theta\to\infty} I(\rho_0)+I(\gamma_0) = \infty$, by working with a subsequence we can assume that $\lim_{\theta\to+\infty}[I(\rho_0)+I(\gamma_0)] = +\infty$. Then the right hand side of (5.5) goes to $-\infty$ since I is the dominating term. Taking $\limsup_{\lambda\to 1+}\limsup_{\epsilon\to 0+}$ on (5.6) and noting $S(\rho) < \infty$, therefore

$$
\overline{f}(\rho) - \underline{f}(\rho) \leqslant 0. \qquad \square
$$

5.1.2. The Cauchy problem

The proof of comparison principle for the Cauchy problem is similar to that of the resolvent problem, we choose to only highlight differences.

Lemma 5.6 *(Comparison principle). Assume that Condition* 1.5 *holds, that $g \in C(\mathcal{P}_2(\mathbb{R}^d))$ satisfies Condition* 1.8*. Let $\overline{U}, \underline{U} : [0,T]\times\mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}\cup\{\pm\infty\}$ be respectively viscosity sub- and super-solution to* (1.46) *satisfying growth estimates*

$$
\overline{U}(t,\rho) \leqslant \zeta_1^*\big(S(\rho)\big) \quad \text{and} \quad -\zeta_2^*\big(S(\rho)\big) \leqslant \underline{U}(t,\rho) \quad \forall (t,\rho)\in[0,T]\times\mathcal{P}_2\big(\mathbb{R}^d\big)
$$

for some $\zeta_1^*, \zeta_2^* : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with sub-linear growth at infinity.

Then

$$
\overline{U}(t,\rho) \leqslant \underline{U}(t,\rho), \quad \text{for all } (t,\rho)\in[0,T]\times\mathcal{P}_2\big(\mathbb{R}^d\big) \text{ satisfying } S(\rho) < \infty.
$$

If in addition $\overline{U}, \underline{U} \in C([0,T]\times\mathcal{P}_2(\mathbb{R}^d))$, then the above inequality holds for all (t,ρ) in $[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$. In particular, there is at most one viscosity solution $U \in C([0,T]\times\mathcal{P}_2(\mathbb{R}^d))$ to (1.46) *satisfying*

$$
\big|U(t,\rho)\big| \leqslant \zeta^*\big(S(\rho)\big) \quad \forall (t,\rho)\in[0,T]\times\mathcal{P}_2\big(\mathbb{R}^d\big)
$$

for some function $\zeta^* : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with sub-linear growth at infinity.

Proof. Let $\lambda > 1$, $c > 0$ and

$$G(t,s;\rho,\gamma) := \lambda\big[\overline{U}(t,\rho) - ct\big] - \underline{U}(s,\gamma) - \frac{\alpha}{2}|t-s|^2 - \frac{\theta}{2}d^2(\rho,\gamma) - \epsilon S(\rho) - \epsilon S(\gamma),$$

where $\alpha, \theta, \epsilon > 0$. Then there exists $(t_0, \rho_0; s_0, \gamma_0) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \times [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ such that

$$G(t_0,s_0;\rho_0,\gamma_0) = \sup_{t,s\in[0,T];\rho,\gamma\in\mathcal{P}_2(\mathbb{R}^d)} G(t,s;\rho,\gamma).$$

It follows that

$$\limsup_{\theta,\alpha\to+\infty}\left[\frac{\alpha}{2}|t_0-s_0|^2 + \frac{\theta}{2}d^2(\rho_0,\gamma_0)\right] = 0.$$

Let ϵ, λ be fixed, $\{(t_0,s_0)\colon \alpha,\theta > 0\}$ is relatively compact in $[0,T]\times[0,T]$ and any limit point has to be of the form $(r_0,r_0) \in [0,T]\times[0,T]$. Also, since

$$\sup_{\alpha,\theta}\big[S(\rho_0) + S(\gamma_0)\big] < \infty,$$

the set $\{(\rho_0,\gamma_0)\colon \alpha,\theta > 0\}$ is relatively compact in $\mathcal{P}_2(\mathbb{R}^d)\times\mathcal{P}_2(\mathbb{R}^d)$ as long as $\epsilon > 0$ is fixed. Hence by working with a subsequence, we assume that

$$\lim_{\alpha,\theta\to\infty} t_0 = \lim_{\alpha,\theta\to\infty} s_0 = r_* = r_*(\epsilon,\lambda), \quad \text{and} \quad \lim_{\alpha,\theta\to\infty}\rho_0 = \lim_{\alpha,\theta\to\infty}\gamma_0 = \rho_* = \rho_*(\epsilon,\lambda).$$

Suppose $r_* > 0$. In this case by taking α, θ large enough if necessary, $t_0, s_0 > 0$. Let

$$U_0(t,\rho) := \lambda ct + \frac{\alpha}{2}|t-s_0|^2 + \frac{\theta}{2}d^2(\rho,\gamma_0) + \epsilon S(\rho),$$
$$U_1(s,\gamma) := -\frac{\alpha}{2}|t_0-s|^2 - \frac{\theta}{2}d^2(\rho_0,\gamma) - \epsilon S(\gamma).$$

Then $(t_0,\rho_0) \in (0,T]\times\mathcal{P}_2(\mathbb{R}^d)$ is a maximum point of $\overline{U} - \lambda^{-1}U_0$, and $(s_0,\gamma_0) \in (0,T]\times\mathcal{P}_2(\mathbb{R}^d)$ is a maximum point of $U_1 - \underline{U}$. Thus by viscosity solution properties,

$$-\lambda c - \alpha(t_0-s_0) + \lambda H\big(\lambda^{-1}U_0\big)(t_0,\rho_0) \geqslant 0$$
$$-\alpha(t_0-s_0) + HU_1(s_0,\gamma_0) \leqslant 0;$$

implying

$$\lambda c \leqslant \lambda H\big(\lambda^{-1}U_0\big)(t_0,\rho_0) - HU_1(s_0,\gamma_0).$$

By exactly the same estimates as in (5.4), the above inequality leads to

$$\lambda c \leqslant -\lambda_\Psi \nu\theta d^2(\rho_0,\gamma_0) - \epsilon\frac{\nu}{2}\big[I(\rho_0) + I(\gamma_0)\big] + (\lambda-1)\|V\vee 0\|_\infty + \big[V(\rho_0) - V(\gamma_0)\big]$$

provided $\epsilon \leqslant \frac{\nu(\lambda-1)}{3\lambda-1}$. Taking $\limsup_{\alpha,\theta\to+\infty}$, by Lemma 2.18,

$$0 < \lambda c \leqslant (\lambda-1)\|V\vee 0\|_\infty.$$

Taking $\lambda_c > 1$ to be such that $(1-\lambda_c^{-1})\|V\vee 0\|_\infty < c/2$, then we have a contradiction $0 < c < c/2$, whenever $1 < \lambda < \lambda_c$. The above arguments lead us to conclude that for ϵ, λ fixed but satisfying $1 < \lambda < \lambda_c$ and $\epsilon \leqslant \frac{\nu(\lambda-1)}{3\lambda-1}$, $r_* = 0$.

Let $r_* = 0$. By the initial condition in the definition of viscosity solution, for each $(t,\rho) \in [0,T]\times\mathcal{P}_2(\mathbb{R}^d)$ satisfying $S(\rho) < \infty$,

$$\begin{aligned}
&\lambda\big[\overline{U}(t,\rho) - ct\big] - \underline{U}(t,\rho) - 2\epsilon S(\rho) \\
&\quad = G(t,t;\rho,\rho) \leqslant \liminf_{\alpha,\theta\to\infty} G(t_0,s_0;\rho_0,\gamma_0) \leqslant \liminf_{\alpha,\theta\to\infty}\big[\lambda\overline{U}(t_0,\rho_0) - \underline{U}(s_0,\gamma_0)\big] \\
&\quad \leqslant \lambda\liminf_{\alpha,\theta\to\infty}\overline{U}(t_0,\rho_0) - \liminf_{\alpha,\theta\to\infty}\underline{U}(s_0,\gamma_0) \leqslant \lambda g(\rho_*) - g(\rho_*) \leqslant (\lambda-1)\|g\vee 0\|_\infty.
\end{aligned}$$

Taking $\lim_{c\to 0+}\lim_{\lambda\to 1+}\lim_{\epsilon\to 0+}$, we obtain $\overline{U}(t,\rho) - \underline{U}(t,\rho) \leqslant 0$ as desired. □

5.2. Existence of viscosity solutions

We show that the value functions U in (1.47) and f in (1.48) are respectively solution to the Cauchy problem (1.46) and the resolvent problem (1.45).

5.2.1. The Cauchy problem

Let $M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$ denote the space of measurable functions from $\mathcal{P}_2(\mathbb{R}^d)$ to $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\}$, which are bounded from above. We define, for $t \geqslant 0$ and $v \in M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$,

$$\begin{aligned} T(t)v(\rho_0) &:= \sup\left\{ v\big(\rho(t)\big) - \int_0^t L(\rho, \dot\rho)\, ds:\ \rho(0) = \rho_0, \rho(\cdot) \in C\big([0,\infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big)\right\} \\ &= \sup\big\{ v(\rho_1) - D(\rho_1, \rho_0; t):\ \rho_1 \in \mathcal{P}_2\big(\mathbb{R}^d\big)\big\}. \end{aligned} \tag{5.7}$$

Then $T(t) : M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}}) \mapsto M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$. Notice that the U in (1.47) is nothing but $U(t, \rho_0) = T(t)g(\rho_0)$.

Lemma 5.7. *For $t, s \geqslant 0$, we have*

$$T(s)T(t)v = T(t+s)v \quad \textit{for all } v \in M^u\big(\mathcal{P}_2\big(\mathbb{R}^d\big); \overline{\mathbb{R}}\big).$$

Proof. Let $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\epsilon > 0$.

There exists $\sigma(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ with $\sigma(0) = \rho_0$ such that

$$T(t+s)v(\rho_0) \leqslant \epsilon + v\big(\sigma(t+s)\big) - \int_0^{t+s} L\big(\sigma(r), \dot\sigma(r)\big)\, dr.$$

Define a new trajectory $\sigma_s(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ by a time shift $\sigma_s(r) := \sigma(s + r)$ for $r \geqslant 0$. Then $\sigma_s(0) = \sigma(s)$ and it follows that

$$\begin{aligned} T(t+s)v(\rho_0) &\leqslant \epsilon + T(t)v\big(\sigma_s(0)\big) - \int_0^s L\big(\sigma(r), \dot\sigma(r)\big)\, dr \\ &\leqslant \epsilon + T(s)T(t)v(\rho_0). \end{aligned} \tag{5.8}$$

On the other hand, there exist $\sigma_i(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$, $i = 1, 2$, satisfying $\sigma_1(0) = \rho_0$ and $\sigma_2(0) = \sigma_1(s)$ such that

$$T(s)T(t)v(\rho_0) \leqslant \epsilon + T(t)v\big(\sigma_1(s)\big) - \int_0^s L(\sigma_1, \dot\sigma_1)\, dr,$$

and

$$T(t)v\big(\sigma_1(s)\big) \leqslant \epsilon + v\big(\sigma_2(t)\big) - \int_0^t L(\sigma_2, \dot\sigma_2)\, dr.$$

Letting $\sigma(r) := \sigma_1(r)$ when $0 \leqslant r \leqslant s$ and $\sigma(r) := \sigma_2(r - s)$ when $r \geqslant s$. Then

$$T(s)T(t)v(\rho_0) \leqslant 2\epsilon + v\big(\sigma(t+s)\big) - \int_0^{t+s} L(\sigma, \dot\sigma)\, dr \leqslant 2\epsilon + T(t+s)v(\rho_0).$$

This together with (5.8) yields $T(s)T(t)v(\rho_0) - 2\epsilon \leqslant T(t+s)v(\rho_0) \leqslant \epsilon + T(s)T(t)v(\rho_0)$ for every $\epsilon > 0$. Hence the lemma follows. $\square$

By Lemma 5.7, $\{T(t)\colon t \geqslant 0\}$ forms a semigroup on $M^u(\mathcal{P}_2(\mathbb{R}^d); \bar{\mathbb{R}})$. We next prove some properties of this semigroup. First, we introduce a class of localization functions $\varphi = \varphi_{L,M} \in C^2(\mathbb{R})$, where

$$\begin{aligned}
&\varphi(r) = r \quad \text{when } r < L, \quad \varphi(r) = L + 1 \quad \text{when } r > L + M,\\
&0 \leqslant \varphi' \leqslant 1 \quad \text{and} \quad \varphi'(r) > 0 \quad \text{if } r < L + M,\\
&\varphi'' \leqslant 0.
\end{aligned}$$

To motivate the utility of such functions φ, we note that $f_0 \notin M^u(\mathcal{P}_2(\mathbb{R}^d); \bar{\mathbb{R}})$ but $\varphi \circ f_0 \in M^u(\mathcal{P}_2(\mathbb{R}^d); \bar{\mathbb{R}})$, where $f_0 \in D_0$.

Proposition 5.8. *Under Condition* 1.5, *the semigroup* $\{T(t)\colon t \geqslant 0\}$ *has the following properties*:

(1) *commuting with addition of a constant, i.e.,* $T(t)(v + c) = T(t)v + c$;
(2) *order preserving, i.e.,* $T(t)v \leqslant T(t)w$ *whenever* $v \leqslant w$;
(3) *for each* $f_1 \in D_1$ *and* $\gamma_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *such that* $S(\gamma_0) < \infty$,

$$\liminf_{t\to 0+} \frac{T(t)f_1(\gamma_0) - f_1(\gamma_0)}{t} \geqslant Hf_1(\gamma_0); \tag{5.9}$$

(4) *for each* $f_0 \in D_0$, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *such that* $S(\rho_0) < \infty$, *for* $L > f_0(\rho_0)$, *and* M *sufficiently large, take* $\varphi = \varphi_{L,M}$, *then*

$$\limsup_{t\to 0+} \frac{T(t)(\varphi \circ f_0)(\rho_0) - \varphi \circ f_0(\rho_0)}{t} \leqslant Hf_0(\rho_0). \tag{5.10}$$

Proof. We only prove (5.9) and (5.10). The others follow directly from the definition.

Let $f_1(\gamma) := -\frac{\theta}{2} d^2(\rho, \gamma) - \epsilon S(\gamma)$, where $\theta, \epsilon > 0$ and $\rho \in \mathcal{P}_2(\mathbb{R}^d)$. Let $\gamma_0 \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $S(\gamma_0) < \infty$. For each $p \in C_c^\infty(\mathbb{R}^d)$, let σ be the path satisfying heat equation

$$\partial_t \sigma = -\nu \operatorname{grad} S(\sigma) - \operatorname{div}(\sigma \nabla p) = \nu \Delta \sigma + \operatorname{div}\big(\sigma \nabla(\nu \Psi - p)\big), \quad \sigma(0) = \gamma_0.$$

Well posedness of the above equation follows from standard parabolic theory.

By Lemma 2.1,

$$\sup_{0\leqslant r\leqslant T} S\big(\sigma(r)\big) + \int_0^T I\big(\sigma(r)\big)\, dr < \infty \quad \text{and} \quad \sigma(\cdot) \in AC^2\big(0, T; \mathcal{P}_2\big(\mathbb{R}^d\big)\big).$$

In particular, we can apply the chain rule formula as explained on p. 233 of [2] to obtain

$$\begin{aligned}
f_1\big(\sigma(t)\big) - f_1(\gamma_0) &= \int_0^t \big\langle \operatorname{grad} f_1(\sigma), \dot{\sigma} \big\rangle_{-1,\sigma}\, dr\\
&= \int_0^t \big\langle \operatorname{grad} f_1(\sigma), -\nu \operatorname{grad} S(\sigma) - \operatorname{div}(\sigma \nabla p) \big\rangle_{-1,\sigma}\, dr.
\end{aligned}$$

Consequently,

$$\begin{aligned}
T(t)f_1(\gamma_0) - f_1(\gamma_0) &= \sup\left\{ f_1\big(\rho(t)\big) - f_1(\gamma_0) - \int_0^t L(\rho, \dot{\rho})\, dr\colon \rho(\cdot) \in C\big([0,\infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big), \rho(0) = \gamma_0 \right\}\\
&\geqslant f_1\big(\sigma(t)\big) - f_1(\gamma_0) - \int_0^t \left(\frac{1}{2} \big\| \dot{\sigma} + \nu \operatorname{grad} S(\sigma) \big\|_{-1,\sigma}^2 - V(\sigma) \right) dr
\end{aligned}$$

$$= \int_0^t \left(\langle \operatorname{grad} f_1(\sigma), -\nu \operatorname{grad} S(\sigma) - \operatorname{div}(\sigma \nabla p) \rangle_{-1,\sigma} - \frac{1}{2} \int_{\mathbb{R}^d} |\nabla p|^2 \, d\sigma + V(\sigma) \right) dr$$

$$= \int_0^t \left[G(\sigma(r)) + \epsilon \nu I(\sigma(r)) + V(\sigma(r)) \right] dr,$$

where

$$G(\sigma) := \left\langle -\theta \operatorname{grad}_\sigma \frac{1}{2} d^2(\rho, \sigma), -\nu \operatorname{grad} S(\sigma) - \operatorname{div}(\sigma \nabla p) \right\rangle_{-1,\sigma} + \epsilon \langle \operatorname{grad} S(\sigma), \operatorname{div}(\sigma \nabla p) \rangle_{-1,\sigma} - \frac{1}{2} \int_{\mathbb{R}^d} |\nabla p|^2 \, d\sigma.$$

By (1.41) and Lemma 2.18, $|G(\sigma)| + |V(\sigma)| \leqslant \zeta_C(I(\sigma))$ in every balls with finite radius

$$B(\rho, C) := \{ \sigma \colon d(\sigma, \rho) \leqslant C \}, \quad C \in \mathbb{R},$$

where $\zeta_C : \mathbb{R} \mapsto \mathbb{R}$ is a sub-linear function possibly depending on C. G is continuous on finite level sets of I (see the arguments in the proof of Lemma 5.1). Apply Lemma B.1, $G + V + \epsilon \nu I$ is lower semicontinuous and bounded from below on every such $B(\rho, C)$. Hence

$$\liminf_{t \to 0+} t^{-1} \left[T(t) f_1(\gamma_0) - f_1(\gamma_0) \right] \geqslant G(\gamma_0) + V(\gamma_0) + \epsilon \nu I(\gamma_0)$$
$$= \langle \operatorname{grad} f_1(\gamma_0), -\nu \operatorname{grad} S(\gamma_0) \rangle_{-1,\gamma_0} + \langle \operatorname{grad} f_1(\gamma_0), p \rangle - \frac{1}{2} \int_{\mathbb{R}^d} |\nabla p|^2 \, d\gamma_0 + V(\gamma_0),$$

where we used the fact that $\langle \operatorname{grad} f_1(\gamma_0), -\operatorname{div}(\gamma_0 \nabla p) \rangle_{-1,\gamma_0} = \langle \operatorname{grad} f_1(\gamma_0), p \rangle$, a consequence of (D.45) in Lemma D.34 of [14]. Therefore, (5.9) follows by taking supremum with respect to $p \in C_c^\infty(\mathbb{R}^d)$ on both side of the above inequality and using the variational definition of $\| \cdot \|_{-1,\gamma_0}$ in (1.18).

We next prove (5.10).

Let $\eta > 0$ be fixed, f_0, ρ_0 and $\varphi = \varphi_{L,M}$ be as given with $L > f(\rho_0)$ and $M > 0$ is arbitrary but fixed for now. For each $t > 0$, there exists a path $\sigma_t(\cdot) = \sigma_{\eta,t,L,M}(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ with $\sigma_t(0) = \rho_0$ such that

$$T(t)(\varphi \circ f_0)(\rho_0) \leqslant \eta t + \varphi \circ f_0(\sigma_t(t)) - \int_0^t L(\sigma_t(r), \dot{\sigma}_t(r)) \, dr. \tag{5.11}$$

Define

$$K_\delta = K_{\delta,L,M} := \{ \rho \in \mathcal{P}_2(\mathbb{R}^d) \colon \varphi' \circ f_0(\rho) \geqslant \delta \}.$$

By Lemma 5.9, there exists some $0 < \delta < 1$ independent of t, such that when M is large enough and t is small enough, $\sigma_t(r) \in K_\delta$ for every $0 \leqslant r \leqslant t$. f_0 is lower semicontinuous and φ' is non-increasing, $\varphi' \circ f_0$ is upper semicontinuous. That is, K_δ is a closed set. Moreover, $\rho \in K_\delta$ implies $f_0(\rho) \leqslant L + M$. Since f_0 has compact finite level sets, K_δ has to be a compact subset in $\mathcal{P}_2(\mathbb{R}^d)$.

Also by Lemma 5.9, the following estimates hold

$$\int_0^t I(\sigma_t(r)) \, dr \leqslant C_{\nu,L,\rho_0} \quad \text{and} \quad \int_0^t \| \dot{\sigma}_t(r) \|^2_{-1,\sigma_t(r)} \, dr \leqslant C_{\nu,L,\rho_0}. \tag{5.12}$$

Therefore, $\sigma_t(\cdot) \in AC^2(0,t; \mathcal{P}(\mathbb{R}^d))$. Denote $m_t(r) := \dot{\sigma}_t(r) + \nu \operatorname{grad} S(\sigma_t(r))$. By the chain rule (p. 233 of [2]) applied to $\varphi \circ f_0$ and the fact that $0 \leqslant \varphi' \leqslant 1$, (5.11) gives

$$
\begin{aligned}
&T(t)(\varphi\circ f_0)(\rho_0)-\varphi\circ f_0(\rho_0)\\
&\quad\leqslant \eta t+\varphi\circ f_0\big(\sigma_t(t)\big)-\varphi\circ f_0\big(\sigma_t(0)\big)-\int_0^t L\big(\sigma_t(r),\dot\sigma_t(r)\big)\,dr\\
&\quad=\eta t+\int_0^t\Big[\varphi'\circ f_0\big(\sigma_t(r)\big)\big\langle \operatorname{grad} f_0\big(\sigma_t(r)\big),\dot\sigma_t(r)\big\rangle_{-1,\sigma_t(r)}-\frac12\big\|m_t(r)\big\|^2_{-1,\sigma_t(r)}+V(\sigma_t(r))\Big]\,dr\\
&\quad\leqslant \eta t+\int_0^t\Big[\varphi'\circ f_0\big(\sigma_t(r)\big)\big\langle \operatorname{grad} f_0\big(\sigma_t(r)\big),-\nu\operatorname{grad} S\big(\sigma_t(r)\big)\big\rangle_{-1,\sigma_t(r)}\\
&\qquad+\varphi'\circ f_0\big(\sigma_t(r)\big)\Big(\big\langle \operatorname{grad} f_0\big(\sigma_t(r)\big),m_t(r)\big\rangle_{-1,\sigma_t(r)}-\frac12\big\|m_t(r)\big\|^2_{-1,\sigma_t(r)}\Big)+V\big(\sigma_t(r)\big)\Big]\,dr\\
&\quad\leqslant \eta t+\int_0^t\big[\varphi'\circ f_0\big(\sigma_t(r)\big)Hf_0\big(\sigma_t(r)\big)+\big(1-\varphi'\circ f_0\big(\sigma_t(r)\big)\big)\|V\vee 0\|_\infty\big]\,dr\\
&\quad=\eta t+t\int_{K_\delta}\big[\big(\varphi'\circ f_0\big)(\sigma)Hf_0(\sigma)+\|V\vee 0\|_\infty\big(1-\varphi'\circ f_0(\sigma)\big)\big]\mu_t(d\sigma),
\end{aligned}
\tag{5.13}
$$

where occupation measure

$$
\mu_t(A):=\frac1t\int_0^t\chi_A\big(\sigma_t(r)\big)\,dr,\quad \forall \text{ Borel set } A\subset\mathcal{P}_2\big(\mathbb{R}^d\big).
$$

By Lemma 5.9, μ_t is supported on compact set K_δ and $d(\sigma_t(r),\rho_0)\leqslant C\sqrt{r}$. Therefore, for each bounded continuous function $G:\mathcal{P}_2(\mathbb{R}^d)\mapsto\mathbb{R}$,

$$
\lim_{t\to0^+}\int_{\mathcal{P}_2(\mathbb{R}^d)}G(\rho)\mu_t(d\rho)=\lim_{t\to0^+}\frac1t\int_0^t G\big(\sigma_t(r)\big)\,dr=G(\rho_0).
$$

Consequently, $\mu_t\Rightarrow\mu_0:=\delta_{\rho_0}$ in the weak convergence of probability measures topology in $\mathcal{P}(\mathcal{P}_2(\mathbb{R}^d))$. This together with the upper semicontinuity of $(\varphi'\circ f_0)Hf_0+\|V\vee 0\|_\infty(1-\varphi'\circ f_0)$ (Lemma 5.10) yields

$$
\begin{aligned}
&\limsup_{t\to0^+}\int_{K_\delta}\big[\big(\varphi'\circ f_0\big)Hf_0+\|V\vee 0\|_\infty\big(1-\varphi'\circ f_0\big)\big]\,d\mu_t\\
&\quad\leqslant\int_{K_\delta}\big[\big(\varphi'\circ f_0\big)Hf_0+\|V\vee 0\|_\infty\big(1-\varphi'\circ f_0\big)\big]\,d\mu_0=Hf_0(\rho_0),
\end{aligned}
$$

where the last equality above follows from $f_0(\rho_0)<L$ and $\varphi'\circ f_0(\rho_0)=1$.

Using the above estimates and taking $\limsup_{t\to0+}t^{-1}$ on both sides of (5.13) leads to (5.10). □

The above proof relies on the following two regularity results.

Lemma 5.9. *Let $\sigma_t(\cdot)$ be the path satisfying* (5.11). *Then there exists some $0<\delta<1$ independent of t and M, such that when M is large and t is small enough, $\sigma_t(r)\in K_\delta$ for $0\leqslant r\leqslant t$. Moreover, $d(\sigma_t(r),\rho_0)\leqslant C\sqrt{r}$ for every $0\leqslant r\leqslant t$, and* (5.12) *holds.*

Proof. Let $\hat\sigma(\cdot)\in C([0,\infty);\mathcal{P}_2(\mathbb{R}^d))$ be a path satisfying

$$
\partial_t\hat\sigma=-\nu\operatorname{grad} S(\hat\sigma)\quad\text{and}\quad\hat\sigma(0)=\rho_0.
$$

By Lemma 2.18, $V(\rho) \geqslant -\hat{\zeta}(I(\rho))$. Together with the conclusion of Lemma 2.5,

$$\int_0^t V\big(\hat{\sigma}(r)\big)\,dr \geqslant -\int_0^t I\big(\hat{\sigma}(r)\big)\,dr - C_0 t \geqslant -\frac{2}{\nu} S(\rho_0) - C_0 t,$$

for some constant $C_0 > 0$. As a consequence,

$$\begin{aligned} T(t)(\varphi \circ f_0)(\rho_0) &\geqslant \varphi \circ f_0\big(\hat{\sigma}(t)\big) - \int_0^t L(\hat{\sigma}, \dot{\hat{\sigma}})\,dr = \varphi \circ f_0\big(\hat{\sigma}(t)\big) + \int_0^t V(\hat{\sigma})\,dr \\ &\geqslant \inf_\rho f_0(\rho) - \frac{2}{\nu} S(\rho_0) - C_0 t. \end{aligned}$$

This together with (5.11) implies in particular that

$$\frac{1}{2}\int_0^t \big\|m_t(r)\big\|^2_{-1,\sigma_t(r)}\,dr \leqslant (\eta + C_0)t + (L+1) - \inf_\rho f_0(\rho) + \frac{2}{\nu} S(\rho_0) + t \sup_\rho V(\rho),$$

where $m_t(r) := \dot{\sigma}_t(r) + \nu \operatorname{grad} S(\sigma_t(r))$. Therefore, by Lemmas 2.5 and 2.6, (5.12) holds. Consequently, $\sigma_t(\cdot) \in AC^2([0,t]; \mathcal{P}_2(\mathbb{R}^d))$, and satisfies

$$d\big(\sigma_t(r), \rho_0\big) \leqslant \int_0^r \big\|\dot{\sigma}_t(s)\big\|_{-1,\sigma_t(s)}\,ds \leqslant C_{\nu,L,\rho_0}\sqrt{r} \quad \text{for all } 0 \leqslant r \leqslant t.$$

Notice that the constant C_{ν,L,ρ_0} is independent of $t \leqslant 1$ and M.

Using absolute continuity of $\sigma_t(\cdot)$ and (5.12), apply the chain rule, for all $t > 0$ sufficiently small and $0 \leqslant r \leqslant t$

$$\begin{aligned} f_0\big(\sigma_t(r)\big) - f_0(\rho_0) &= f_0\big(\sigma_t(r)\big) - f_0\big(\sigma_t(0)\big) \\ &= \frac{\theta}{2}\big[d^2\big(\sigma_t(r), \gamma\big) - d^2(\rho_0, \gamma)\big] + \epsilon \int_0^r \big\langle \operatorname{grad} S\big(\sigma_t(s)\big), \dot{\sigma}_t(s)\big\rangle_{-1,\sigma_t(s)}\,ds \\ &\leqslant \frac{\theta}{2} d\big(\sigma_t(r), \rho_0\big)\big[d\big(\sigma_t(r), \rho_0\big) + 2d(\rho_0, \gamma)\big] + \epsilon \int_0^r \sqrt{I\big(\sigma_t(s)\big)}\big\|\dot{\sigma}_t(s)\big\|_{-1,\sigma_t(s)}\,ds \\ &\leqslant C_{\epsilon,\nu,L,\rho_0}. \end{aligned}$$

The constant $C_{\epsilon,\nu,L,\rho_0}$ is independent of t and M. Therefore, we can select $M > 0$ large enough and t small enough such that there exists some $0 < \delta < 1$,

$$\varphi' \circ f_0\big(\sigma_t(r)\big) \geqslant \delta \quad \forall r \in [0,t].$$

That is, $\sigma_t(r) \in K_\delta$. $\square$

Lemma 5.10. $(\varphi' \circ f_0)Hf_0 + \|V \vee 0\|_\infty(1 - \varphi' \circ f_0)$: $K_\delta \mapsto \mathbb{R} \cup \{-\infty\}$ *is upper semicontinuous and bounded from above.*

Proof. Hf_0 is bounded from above by Lemma 5.1, $0 \leqslant \varphi' \leqslant 1$, therefore

$$\big(\varphi' \circ f_0\big)Hf_0 + \|V \vee 0\|_\infty\big(1 - \varphi' \circ f_0\big)$$

is bounded from above as well.

To prove the upper semicontinuity of $(\varphi' \circ f_0)Hf_0 + \|V \vee 0\|_\infty(1 - \varphi' \circ f_0)$ on K_δ, let $K_\delta \ni \rho_n \to \bar{\rho} \in K_\delta$. Consider a subsequence if necessary, we may assume that

$$-\infty < -C \leqslant \inf_n \big[\big(\varphi' \circ f_0\big)(\rho_n)Hf_0(\rho_n) + \|V \vee 0\|_\infty\big(1 - \varphi' \circ f_0(\rho_n)\big)\big].$$

Since $\varphi' \circ f_0(\rho) \geqslant \delta$ for $\rho \in K_\delta$, $-\infty < \inf_n Hf_0(\rho_n)$. The dominating term in Hf_0 is $-\epsilon\nu I$, therefore $\sup_n I(\rho_n) < \infty$. By Lemma 5.3 and the lower semicontinuity of S, $\lim_{n\to\infty} S(\rho_n) = S(\bar{\rho})$, implying

$$\lim_{n\to\infty} \varphi' \circ f_0(\rho_n) = \varphi' \circ f_0(\bar{\rho}).$$

It follows from this, the upper boundedness and upper semicontinuity of Hf_0 from Lemma 5.1 that

$$\begin{aligned}
&\limsup_{n\to\infty}\big[\varphi' \circ f_0(\rho_n)Hf_0(\rho_n) - \varphi' \circ f_0(\bar{\rho})Hf_0(\bar{\rho})\big] \\
&\quad \leqslant \limsup_{n\to\infty}\big[\varphi' \circ f_0(\rho_n) - \varphi' \circ f_0(\bar{\rho})\big]Hf_0(\rho_n) \\
&\qquad + \varphi' \circ f_0(\bar{\rho}) \limsup_{n\to\infty}\big[Hf_0(\rho_n) - Hf_0(\bar{\rho})\big] \leqslant 0.
\end{aligned}$$

Therefore, the conclusion follows. □

We prove that the value function $U : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}$ defined by (1.47) is a viscosity solution to (1.46). In view of the estimates in Lemmas 4.5, 4.7, 4.8, assumptions of the following lemmas are satisfied at various levels of generality.

Lemma 5.11. *Assume that Condition* 1.5 *holds, that g satisfies Condition* 1.8. *Then U is a viscosity super-solution to* (1.46) *satisfying* (4.11).

Proof. By Lemmas 4.7 and 4.8, $U = U(t, \rho)$ is continuous on $[0, \infty) \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant C\}$ for each $C \in \mathbb{R}$. Estimate (4.11) follows from Lemma 4.5.

Let U_1 be given by (1.63) and suppose that $(s_0, \gamma_0) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ is such that (1.64) holds. Then U_1 is bounded from above, $S(\gamma_0) < \infty$ and

$$U(s, \gamma) - U(s_0, \gamma_0) \geqslant U_1(s, \gamma) - U_1(s_0, \gamma_0), \quad \forall (s, \gamma) \in [0, T] \times \mathcal{P}_2\big(\mathbb{R}^d\big).$$

First, we consider the situation where $s_0 > 0$.

By (5.9) and properties of the semigroup T in Proposition 5.8, for $0 < r < s_0$,

$$\begin{aligned}
0 = U(s_0, \gamma_0) - U(s_0, \gamma_0) &= T(r)\big(U(s_0 - r, \cdot) - U(s_0, \gamma_0)\big)(\gamma_0) \\
&\geqslant T(r)\big(U_1(s_0 - r, \cdot) - U_1(s_0, \gamma_0)\big)(\gamma_0).
\end{aligned} \tag{5.14}$$

On the other hand, because of the special form of U_1,

$$\frac{\alpha}{2}|t - s_0 + r|^2 - \frac{\alpha}{2}|t - s_0|^2 = r\alpha(t - s_0) + \frac{\alpha}{2}r^2 = \frac{\partial}{\partial s}U_1(s_0, \gamma_0)r + \frac{\alpha}{2}r^2,$$

and for $0 < r < s_0$, $\gamma \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned}
U_1(s_0 - r, \gamma) &= U_1(s_0, \gamma) + \frac{\alpha}{2}|t - s_0|^2 - \frac{\alpha}{2}|t - s_0 + r|^2 \\
&= U_1(s_0, \gamma) - r\frac{\partial}{\partial s}U_1(s_0, \gamma_0) - \frac{\alpha}{2}r^2.
\end{aligned}$$

Combined with (5.14), therefore

$$0 \geqslant r^{-1}\big\{T(r)\big(U_1(s_0, \cdot)\big)(\gamma_0) - U_1(s_0, \gamma_0)\big\} - \frac{\partial}{\partial s}U_1(s_0, \gamma_0) - \frac{\alpha}{2}r.$$

The conclusion follows by taking $r \to 0+$ and by (5.9).

Now consider $s_0 = 0$. For every $0 < C < \infty$, by Lemma 4.7, U is lower semicontinuous on $[0,T] \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant C\}$. Therefore

$$\liminf_{t\to 0^+,\gamma'\to\gamma_0,S(\gamma')\leqslant C} U\bigl(t,\gamma'\bigr) \geqslant U(0,\gamma_0) = g(\gamma_0). \qquad \square$$

The case for sub-solution is similar but more complicated because of a localization argument to bound test functions.

Lemma 5.12. *Assume that Condition* 1.5 *holds, that g satisfies Condition* 1.8. *Then the U is a viscosity sub-solution to* (1.46) *satisfying* (4.11).

Proof. Because of Lemmas 4.7 and 4.8, $U = U(t,\rho)$ is continuous on $[0,\infty) \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant C\}$ for each $C \in \mathbb{R}$. By Lemma 4.5, (4.11) holds.

Let U_0 be defined as in (1.60) and $(t_0,\rho_0) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ be such that

$$(U - U_0)(t_0,\rho_0) = \sup_{(t,\rho)\in[0,T]\times\mathcal{P}_2(\mathbb{R}^d)} (U - U_0)(t,\rho).$$

In particular, $S(\rho_0) < \infty$.

We first consider the situation where $t_0 > 0$.

We can rewrite the function U_0 as

$$U_0(t,\rho) = \frac{\alpha}{2}|t-s|^2 + f_0(\rho), \qquad f_0(\rho) := \frac{\theta}{2}d^2(\rho,\gamma) + \epsilon S(\rho) + c.$$

Let

$$L := \|U \vee 0\|_\infty - U(t_0,\rho_0) + U_0(t_0,\rho_0) \vee 0 + 1 > f_0(\rho_0) \vee 0.$$

By Proposition 5.8, if we take $M > 0$ sufficiently large and $\varphi = \varphi_{L,M} \in C^2(\mathbb{R})$ satisfying

$$\begin{aligned} &\varphi(r) = r \quad \text{for } r < L, \qquad \varphi(r) = L+1 \quad \text{for } r > L+M,\\ &\varphi'(r) > 0 \quad \text{for } r < L+M \quad \text{and} \quad \varphi'(r) \leqslant 1, \qquad \varphi'' \leqslant 0, \end{aligned}$$

then (5.10) holds. Now define

$$\tilde{U}_0(t,\rho) := \frac{\alpha}{2}|t-s|^2 + \varphi \circ f_0(\rho) \quad \text{for } (t,\rho) \in [0,T] \times \mathcal{P}_2\bigl(\mathbb{R}^d\bigr).$$

Then whenever $f_0(\rho) > L$, we get $\varphi \circ f_0(\rho) > L$, and hence

$$\begin{aligned} U(t,\rho) - \tilde{U}_0(t,\rho) &\leqslant \|U \vee 0\|_\infty - \frac{\alpha}{2}|t-s|^2 - \varphi \circ f_0(\rho)\\ &\leqslant \|U \vee 0\|_\infty - \frac{\alpha}{2}|t-s|^2 - L\\ &< U(t_0,\rho_0) - U_0(t_0,\rho_0) = U(t_0,\rho_0) - \tilde{U}_0(t_0,\rho_0). \end{aligned}$$

Moreover if $f_0(\rho) \leqslant L$, we have

$$U(t,\rho) - \tilde{U}_0(t,\rho) = U(t,\rho) - U_0(t,\rho) \leqslant U(t_0,\rho_0) - U_0(t_0,\rho_0) = U(t_0,\rho_0) - \tilde{U}_0(t_0,\rho_0).$$

In the last step, we used the fact that $\varphi \circ f_0(\rho_0) = f(\rho_0)$ which follows from the definition of L. In summary,

$$U(t,\rho) - U(t_0,\rho_0) \leqslant \tilde{U}_0(t,\rho) - \tilde{U}_0(t_0,\rho_0) \quad \forall (t,\rho) \in [0,T] \times \mathcal{P}_2\bigl(\mathbb{R}^d\bigr).$$

Therefore by properties of the semigroup T in Proposition 5.8, for $0 < r < t_0$,

$$\begin{aligned} 0 &= U(t_0,\rho_0) - U(t_0,\rho_0) = T(r)\bigl(U(t_0-r,\cdot) - U(t_0,\rho_0)\bigr)(\rho_0)\\ &\leqslant T(r)\bigl(\tilde{U}_0(t_0-r,\cdot) - \tilde{U}_0(t_0,\rho_0)\bigr)(\rho_0). \end{aligned} \tag{5.15}$$

Note that, for $0 < r < t_0$,

$$\frac{\alpha}{2}|t_0 - r - s|^2 - \frac{\alpha}{2}|t_0 - s|^2 = -r\alpha(t_0 - s) + \frac{\alpha}{2}r^2 = -r\frac{\partial}{\partial t}U_0(t_0, \rho_0) + \frac{\alpha}{2}r^2$$

and for every $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\tilde{U}_0(t_0 - r, \rho) - \tilde{U}_0(t_0, \rho_0) = \varphi \circ f_0(\rho) - \varphi \circ f_0(\rho_0) - r\frac{\partial}{\partial t}U_0(t_0, \rho_0) + \frac{\alpha}{2}r^2.$$

This, together with (5.15), gives

$$0 \leqslant r^{-1}\big\{T(r)(\varphi \circ f_0)(\rho_0) - \varphi \circ f_0(\rho_0)\big\} - \frac{\partial}{\partial t}U_0(t_0, \rho_0) + \frac{\alpha}{2}r.$$

Taking $r \to 0+$ and using (5.10), we obtain

$$0 \leqslant Hf_0(\rho_0) - \frac{\partial}{\partial t}U_0(t_0, \rho_0) = HU_0(t_0, \rho_0) - \frac{\partial}{\partial t}U_0(t_0, \rho_0)$$

as desired.

Next, we consider the case of $t_0 = 0$. For each $0 < C < \infty$, by upper semicontinuity of U on $[0, T] \times \{\rho \in \mathcal{P}_2(\mathbb{R}^d)\colon\ S(\rho) \leqslant C\}$ (Lemma 4.8),

$$\limsup_{t\to 0^+, \rho'\to\rho_0, S(\rho')\leqslant C} U\big(t, \rho'\big) \leqslant U(0, \rho_0) = g(\rho_0). \qquad \Box$$

5.2.2. The resolvent equation

For each $v \in M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$ and $\alpha > 0$, we define

$$R_\alpha v(\rho_0) = \sup\left\{ \int_0^\infty e^{-\alpha^{-1}s}\big[\alpha^{-1}v\big(\rho(s)\big) - L\big(\rho(s), \dot{\rho}(s)\big)\big]\,ds\colon \quad \rho(\cdot) \in C\big([0, \infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big),\ \rho(0) = \rho_0 \right\}.$$

It follows then $R_\alpha : M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}}) \mapsto M^u(\mathcal{P}_2(\mathbb{R}^d); \overline{\mathbb{R}})$. Throughout this section, we assume $\|h \vee 0\|_\infty < \infty$. Using the same arguments as in the proof of Lemma 5.7, we have the following:

Lemma 5.13 *(Bellman's principle). Suppose that Condition* 1.5 *holds, that h satisfies Condition* 1.6 *with the V replaced by h. Then the value function f in* (1.48) *satisfies*

$$f(\rho_0) = \sup\left\{ \int_0^t e^{-\alpha^{-1}s}\big[\alpha^{-1}h\big(\rho(s)\big) - L\big(\rho(s), \dot{\rho}(s)\big)\big]ds + e^{-\alpha^{-1}t}f\big(\rho(t)\big)\colon \quad \rho \in C\big([0, \infty); \mathcal{P}_2\big(\mathbb{R}^d\big)\big),\ \rho(0) = \rho_0 \right\}$$

for all $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

We use the above result to prove that the value function f is a viscosity solution to (1.45).

Lemma 5.14. *Suppose that Condition* 1.5 *holds and h satisfies Condition* 1.6 *with the V replaced by h. Then the function f satisfies* (4.2) (*Lemma* 4.1)*, and it is a viscosity super-solution to* (1.45).

Proof. First, f is continuous on finite level sets of S by Lemmas 4.3 and 4.4.

Next, we prove that the viscosity super-solution property is satisfied. Let $f_1 \in D_1$ and $\gamma_0 \in \mathcal{P}_2(\mathbb{R}^d)$ be such that

$$(f_1 - f)(\gamma_0) \geqslant f_1(\gamma) - f(\gamma) \quad \forall \gamma \in \mathcal{P}_2(\mathbb{R}^d).$$

Then $S(\gamma_0) < \infty$ and $f_1(\gamma) - f_1(\gamma_0) \leqslant f(\gamma) - f(\gamma_0)$. Hence by Proposition 5.8,

$$\begin{aligned} T(t)f_1(\gamma_0) - f_1(\gamma_0) &= T(t)\big(f_1 - f_1(\gamma_0)\big)(\gamma_0) \\ &\leqslant T(t)\big(f - f(\gamma_0)\big)(\gamma_0) = T(t)f(\gamma_0) - f(\gamma_0). \end{aligned} \tag{5.16}$$

By (5.9),

$$Hf_1(\gamma_0) \leqslant \liminf_{t\to 0^+} t^{-1}\big(T(t)f(\gamma_0) - f(\gamma_0)\big).$$

We claim that

$$\liminf_{t\to 0^+} t^{-1}\big(T(t)f(\gamma_0) - f(\gamma_0)\big) \leqslant \alpha^{-1}(f-h)(\gamma_0), \tag{5.17}$$

hence conclude the proof.

We prove (5.17) next. By Remark 5.4, we only need to prove the case $h = 0$. Let $\eta > 0$. For each $t > 0$, there exists a $\gamma_t(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ satisfying $\gamma_t(0) = \gamma_0$, and

$$T(t)f(\gamma_0) \leqslant \eta t + f\big(\gamma_t(t)\big) - \int_0^t L\big(\gamma_t(s), \dot{\gamma}_t(s)\big)\,ds. \tag{5.18}$$

This together with Lemma 5.13 gives

$$\begin{aligned} \frac{T(t)f(\gamma_0) - f(\gamma_0)}{t} &\leqslant \eta - \frac{1}{t}\int_0^t e^{-\alpha^{-1}s}\alpha^{-1}h\big(\gamma_t(s)\big)\,ds + \frac{1 - e^{-\alpha^{-1}t}}{t} f\big(\gamma_t(t)\big) \\ &\quad + \|V \vee 0\|_\infty \bigg[1 + \alpha\frac{e^{-\alpha^{-1}t} - 1}{t}\bigg]. \end{aligned}$$

Let $m_t(r) := \dot{\gamma}_t(r) + \nu \operatorname{grad} S(\gamma_t(r))$. By (5.18) and $\|f \vee 0\| + \|V \vee 0\| < \infty$, we have

$$\frac{1}{2}\int_0^t \big\|m_t(r)\big\|^2_{-1,\sigma_t(r)}\,dr \leqslant C < \infty, \quad 0 \leqslant t \leqslant 1.$$

Consequently, by Lemmas 2.5 and 2.6,

$$S\big(\gamma_t(t)\big) \leqslant C, \quad \text{and} \quad d\big(\gamma_t(t), \gamma_0\big) \leqslant \int_0^t \big\|\dot{\gamma}_t(r)\big\|_{-1,\gamma_t(r)}\,dr \leqslant C\sqrt{t},$$

where the constant C is independent of $t \leqslant 1$. By continuity of f on level sets of S, (5.17) follows with $h = 0$. □

Lemma 5.15. *Suppose that Condition* 1.5 *holds and h satisfies Condition* 1.6 *with the V replaced by h. Then the function f satisfies* (4.2) (*Lemma* 4.1) *and is a viscosity sub-solution to* (1.45).

Proof. f is continuous on finite level sets of S by Lemmas 4.3 and 4.4. Let $f_0 \in D_0$ and $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ be such that

$$(f - f_0)(\rho_0) \geqslant f(\rho) - f_0(\rho) \quad \forall \rho \in \mathcal{P}_2(\mathbb{R}^d).$$

Then $S(\rho_0) < \infty$. Let

$$L := \|f \vee 0\|_\infty - f(\rho_0) + f_0(\rho_0) \vee 0 + 1 > f_0(\rho_0) \vee 0.$$

By Proposition 5.8, if we take $M > 0$ sufficiently large and $\varphi = \varphi_{L,M} \in C^2(\mathbb{R})$ satisfying

$$\varphi(r) = r \quad \text{for } r < L, \qquad \varphi(r) = L+1 \quad \text{for } r > L+M,$$
$$\varphi'(r) > 0 \quad \text{for } r < L+M \quad \text{and} \quad \varphi'(r) \leqslant 1, \qquad \varphi'' \leqslant 0,$$

then (5.10) holds. Arguing as in Lemma 5.12,

$$f(\rho) - f(\rho_0) \leqslant \varphi \circ f_0(\rho) - \varphi \circ f_0(\rho_0), \quad \forall \rho \in \mathcal{P}_2(\mathbb{R}^d).$$

Hence by Proposition 5.8,

$$\begin{aligned} T(t) f(\rho_0) - f(\rho_0) &= T(t)\big(f - f(\rho_0)\big)(\rho_0) \leqslant T(t)\big(\varphi \circ f_0 - \varphi \circ f_0(\rho_0)\big)(\rho_0) \\ &= T(t)(\varphi \circ f_0)(\rho_0) - \varphi \circ f_0(\rho_0), \end{aligned} \tag{5.19}$$

giving

$$\limsup_{t \to 0^+} t^{-1}\big(T(t) f(\rho_0) - f(\rho_0)\big) \leqslant H f_0(\rho_0).$$

The conclusion follows if we show that

$$\alpha^{-1}(f - h) \leqslant \limsup_{t \to 0^+} t^{-1}\big(T(t) f(\rho_0) - f(\rho_0)\big). \tag{5.20}$$

Again, we only need to deal with the case $h = 0$ (see Remark 5.4).

Let $\eta > 0$. Then for each $t > 0$, by Lemma 5.13 there exists a trajectory $\sigma_t(\cdot) \in C([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ such that $\sigma_t(0) = \rho_0$, and (take $h = 0$)

$$f(\rho_0) \leqslant \eta t - \int_0^t e^{-\alpha^{-1} s} L\big(\sigma_t(s), \dot{\sigma}_t(s)\big)\, ds + e^{-\alpha^{-1} t} f\big(\sigma_t(t)\big). \tag{5.21}$$

This implies that

$$e^{-\alpha^{-1} t} T(t) f(\rho_0) \geqslant f(\rho_0) - \eta t + \int_0^t \big(e^{-\alpha^{-1} s} - e^{-\alpha^{-1} t}\big) L\big(\sigma_t(s), \dot{\sigma}_t(s)\big)\, ds.$$

Since $L(\sigma_t(s), \dot{\sigma}_t(s)) \geqslant -\|V \vee 0\|_\infty$, it follows that

$$\frac{e^{\alpha^{-1} t} - 1}{t} f(\rho_0) - \eta e^{\alpha^{-1} t} + \|V \vee 0\|_\infty \left[1 + \alpha \frac{1 - e^{\alpha^{-1} t}}{t}\right] \leqslant \frac{T(t) f(\rho_0) - f(\rho_0)}{t}.$$

Taking $t \to 0^+$ on both side, (5.20) follows with $h = 0$. □

Appendix A. Variational principle, relaxed formulation and probabilities – the heuristic ideas

The stochastic connections showing up in some proofs are more than a coincidence. Next, we expose origin of the stochastic arguments in a heuristic way. The key is to view minimization of action functional (1.23) from a relaxed point of view, which helps to explain the seemingly ad hoc addition of $+\nu \operatorname{grad} S(\rho)$ in the kinetic energy (1.21) term in a natural way.

To highlight ideas, we will first re-examine finite dimensional, deterministic variational principle in classical point mechanics as limit of another class of variational principles defined on probability measures over path space. Recall that there is a natural embedding of path space into probability measures over path space through identity map $x_0(\cdot) \mapsto \delta_{x_0(\cdot)}(dx(\cdot))$, we now minimize over measures instead of paths. The variational selection of the "most probable" measure has an interpretation of path-space-entropy-minimization which can be viewed as second law of thermodynamic defined on path space. After the finite dimensional point mechanic situation is cleared, we make analogy to continuum mechanics situation by viewing it as an infinite particle limit. Direct computation reveals that the term $\nu \operatorname{grad} S$ has to be included in kinetic energy, accounting for the indistinguishability/exchangeability of the particles, once we take the relaxed formulation to start with.

A.1. Relaxation of variational problems

Two abstract but simple variational formulas form the theoretical foundation of what follows. Let S be a metric space, then

(1) Entropy–pressure formula: for $f \in C_b(S)$,

$$\log \int_S e^{f(x)} P(dx) = \sup_{Q \in \mathcal{P}(S)} \left\{ \int_S f\,dQ - R(Q \| P) \right\} = -\inf_Q \left\{ R\left(Q \| P^f\right) \right\} + \log Z^f \tag{A.1}$$

where renormalized measure $dP^f := (Z^f)^{-1} e^f dP$ with Z^f being the normalization constant.

(2) Laplace principle: for a large class of g and $\mu \in \mathcal{P}(S)$ (assuming $g(x) = +\infty$ when $x \notin \operatorname{supp}(\mu)$),

$$\lim_{\epsilon \to 0+} -\epsilon \log \int_S e^{-\epsilon^{-1} g(x)} \mu(dx) = \inf_{x \in S} g(x). \tag{A.2}$$

In the above, to distinguish entropies defined at different levels (e.g. path space vs. state space), we now use R instead of S to denote abstract relative entropy.

Suppose that $P = P_\epsilon$ has ϵ dependency with limiting behavior $dP_\epsilon \stackrel{\sim}{=} Z_\epsilon^{-1} e^{-\epsilon^{-1} I(x)}\, d\mu$ (i.e. large deviation), then

$$\begin{aligned}
\inf_{x \in S} \left\{ I(x) - V(x) \right\} &= \lim_{\epsilon \to 0+} -\epsilon \log \int_S e^{\epsilon^{-1}(V(x) - I(x))} \mu(dx) \\
&= -\lim_{\epsilon \to 0} \epsilon \log \int_S e^{\epsilon^{-1} V(x)} P_\epsilon(dx) - \epsilon \log Z_\epsilon = \lim_{\epsilon \to 0} \inf_Q \epsilon R\left(Q \| P_\epsilon^V\right).
\end{aligned} \tag{A.3}$$

On the left, we have a variational problem which is not clear to have a unique solution. On the right, before taking the limit, we have a convex optimization problem in space of measures, it always has a unique solution.

The heuristic discussions above can be made rigorous using the language of large deviation in probability theory. Such theory was originally formulated by Donsker–Varadhan, Freidlin–Wentzell, among others. For an exposition usefully stated in terms of variational problems and Hamilton–Jacobi–Bellman equations, see Feng and Kurtz [14]. See also there for an extensive list of references.

A.2. Classical point mechanics and probability theory

We first recall that a d-dimensional standard Brownian motion W (on a finite time interval) can be viewed as a metric space $S := C([0,T]; \mathbb{R}^d)$-valued random variable. Let $\nu > 0$, the probability law $P \in \mathcal{P}(S)$ of time-rescaled Brownian motion $X_\nu(t) = x_0 + W(\nu t)$ is informally speaking

$$P_\nu\big(dx(\cdot)\big) := P\big(X_\nu(\cdot) \in dx(\cdot)\big) = Z_\nu^{-1} e^{-\nu^{-1} \frac{1}{2} \int_0^T |\dot{x}|^2 ds} \pi\big(dx(\cdot)\big) \tag{A.4}$$

where Z_ν is a normalizing constant and π is some kind of "volume measure" on the path space S. Our first important observation is that the term on the exponent exactly matches with kinetic energy of a classical particle.

Adding a potential energy V term, and a penalization function f realizing the terminal condition of position of the path at time T, we arrive at an un-normalized (as opposed to probability) measure on path space

$$\begin{aligned}
&Z_\nu^{-1} e^{-\nu^{-1}\{-f(x(T)) + \int_0^T L(x,\dot{x})\,ds\}} \pi\big(dx(\cdot)\big) \\
&\quad = e^{\nu^{-1}\{f(x(T)) + \int_0^T V(x(s))\}} \Big(Z_\nu^{-1} e^{-\nu^{-1} \frac{1}{2} \int_0^T |\dot{x}|^2 ds} \pi\big(dx(\cdot)\big)\Big) \\
&\quad = e^{\nu^{-1}\{f(x(T)) + \int_0^T V(x(s))\,ds\}} P_\nu\big(dx(\cdot)\big),
\end{aligned} \tag{A.5}$$

where $L(x,p) := \frac{1}{2}|p|^2 - V(x)$ is Lagrangian. Although it is not clear how to make sense out of the volume measure π in (A.4), we note that the very right hand side of the last line above always makes perfect sense and is rigorously well defined – it is just the probability measure of a Brownian motion with initial value x_0. If we choose

$$f(x) = -\infty\chi(x \neq x_1) + 0\chi(x = x_1)$$

then we are forcing the measure to be only supported on those path with initial value $x(0) = x_0$ and terminal value $x(T) = x_1$. Such measure is known as a Brownian bridge. If we do not want to use the bridge process, we can choose another layer of approximation by approximating such bad test function f by sequence of smooth ones. To streamline presentation, we do not distinguish the bad f and its smooth approximations in what follows. Indeed, one can even think of everything in the absence of f, provided every probability measure is replaced by a probability measure induced by a bridge process.

Following derivation of (A.3), first by Laplace principle (A.2),

$$\begin{aligned}
&\lim_{\nu\to 0+} -\nu \log E\big[e^{\nu^{-1}\{f(X_\nu(T))+\int_0^T V(X_\nu(s))\,ds\}}\big] \\
&\quad = \lim_{\nu\to 0+} -\nu \log \int_{x(\cdot)\in S} Z_\nu^{-1} e^{\nu^{-1}\{f(x(T))-\int_0^T L(x,\dot{x})\,ds\}}\pi\big(dx(\cdot)\big) \\
&\quad = \inf\left\{-f\big(x(T)\big) + \int_0^T L\big(x(s),\dot{x}(s)\big)\,ds\colon\ x(\cdot)\in S\right\} \\
&\quad = \inf\left\{\int_0^T L(x,\dot{x})\,ds\colon\ x(0) = x_0, x(T) = x_1, x(\cdot)\in S\right\}.
\end{aligned} \tag{A.6}$$

Then, by entropy formula (A.1),

$$\begin{aligned}
&-\nu \log E\big[e^{\nu^{-1}\{f(X_\nu(T))+\int_0^T V(X_\nu(s))\,ds\}}\big] \\
&\quad = \inf\left\{-E^Q\left[f\big(X_\nu(T)\big) + \int_0^T V\big(X_\nu(s)\big)\,ds\right] + \nu R(Q\|P_\nu)\colon\ Q\in\mathcal{P}\big(C_{x_0,x_1}\big([0,T];\mathbb{R}^d\big)\big)\right\} \\
&\quad = \inf\big\{\nu R\big(Q\|P_\nu^V\big)\colon\ Q\in\mathcal{P}\big(C_{x_0,x_1}\big([0,T];\mathbb{R}^d\big)\big)\big\} - \nu\log Z_{\nu,V}(1),
\end{aligned} \tag{A.7}$$

where $C_{x_0,x_1}([0,T];\mathbb{R}^d)$ is the collection of paths in $C([0,T];\mathbb{R}^d)$ with initial value $x(0) = x_0$ and terminal value $x(T) = x_1$. The new probability measure P_ν^V is re-normalized version of (A.5) to take care of potential V and terminal constraint $x(T) = x_1$:

$$P_\nu^V(\varphi) = Z_{\nu,V}^{-1}(1)Z_{\nu,V}(\varphi), \qquad Z_{\nu,V}(\varphi) = E\big[\varphi\big(X_\nu(\cdot)\big)e^{\nu^{-1}\{f(X_\nu(T))+\int_0^T V(X_\nu(s))\,ds\}}\big], \quad \varphi\in B(S).$$

By convexity of relative entropy R, the minimization problem in (A.7) has one and *only one* solution P_ν^V.

Let us record an important message that the above calculations give us: using the natural embedding $C([0,T],\mathbb{R}^d) \hookrightarrow \mathcal{P}(C([0,T];\mathbb{R}^d))$ through $x(\cdot)\mapsto\delta_{x(\cdot)}$, we regulated (by introducing small parameter $\nu > 0$) and relaxed (in the sense of Young measures) the fully nonlinear problem of path space action minimization into a well-posed linear space convex minimization problem.

Turning to PDE connection to classical mechanics, let

$$H(x,p) = \frac{1}{2}|p|^2 + V(x)$$

be the Legendre transform of L. Then the minimization problem

$$u(t,x_0) := \inf\left\{f\big(x(t)\big) + \int_0^t L(x,\dot{x})\,ds\colon\ x(0) = x_0, x\in C\big([0,\infty);\mathbb{R}^d\big)\right\}$$

solves classical Hamilton–Jacobi equation

$$\partial_t u + H(x,\nabla_x u) = 0, \quad u(0,x) = f(x). \tag{A.8}$$

Historically in classical mechanics, u is used to construct canonical transformation to solve Hamiltonian ODE which is action minimizer $x(t)$ defined using the Lagrangian L. Suppose that u is smooth, then

$$\dot{x}(t) = \nabla u\big(t, x(t)\big). \tag{A.9}$$

Before passing ν to the limit, for each ν fixed,

$$u_\nu(t, x_0) = -\nu \log E\big[e^{\nu^{-1}\{f(X_\nu(T))+\int_0^T V(X_\nu(s))\,ds\}}\big|X_\nu(0) = x_0\big]$$

solves a viscous version of (A.8)

$$\partial_t u_\nu + H_\nu\big(x, \nabla_x u_\nu, D^2 u_\nu\big) = 0, \quad u_\nu(0, x) = f(x), \tag{A.10}$$

where $D^2 u$ is the Hessian matrix of u and

$$H_\nu(x, p, P) = H(x, p) + \frac{\nu}{2}\mathrm{Tr}P.$$

A.3. Continuum mechanics and the important role of exchangeability of particles

In this case, in addition to ν, we also have an extra parameter n, the number of particles. The role of "small" noise, which was previously played by ν, is also played by n. To simplify, we will take ν fixed and only consider the effect of $n \to \infty$. Indeed, to simplify even more, we will just focus on the pressure-less situation. That is, no internal energy in the potential term. We also assume $\Psi = 0$ and just treat the case formally (one can think of each particle lives in a periodic hyper-cube).

Let $\{X_i(t) := x_i + W_i(\nu t)\colon i = 1, 2, \ldots\}$ be a sequence of independent identically distributed Brownian particles. Repeating derivation of (A.4), the n ordered particles

$$\vec{X}(t) := \big(X_1(t), \ldots, X_n(t)\big)$$

defines a measure on $C([0, \infty); (\mathbb{R}^d)^n)$

$$P\big(d\vec{x}(\cdot)\big) := Z_\nu^{-n} e^{-\frac{1}{2}\nu^{-1}\int_0^T \sum_{i=1}^n |\dot{x}_i|^2\,ds} \pi\big(dx_1(\cdot)\big)\ldots\pi\big(dx_n(\cdot)\big),$$

which is the joint probability of n-copies of independent Brownian motions. Adding potential term

$$V(\vec{x}) = \frac{1}{n}\sum_i \phi(x_i) + \frac{1}{n^2}\frac{1}{2}\sum_{i\neq j}\Phi(x_i - x_j),$$

the pressure function becomes

$$-\nu \log E\big[e^{\nu^{-1}\{f(\vec{x}(T))+\int_0^T V(\vec{x}(t))\,dt\}}\big].$$

However, the quantity which we will pass to the limit is not $\vec{x}$. We are interested in measure-valued process

$$\rho_n(t, dx) = n^{-1}\sum_{i=1}^n \delta_{X_i(t)}(dx),$$

which is a "low-dimensional" functional of $\vec{X}$. In particular, ordering information of $\vec{X}$ is erased in ρ_n by averaging. Interestingly, $\rho_n(t)$ is a probability measure-valued Markov process of its own (that is, it forms a closed system of evolution equations). If we permute the orders of $\vec{X}(t)$ at any time t and run the process for dt amount of time, then we cannot tell the difference by just observing ρ_n – we are averaging over $n!$ different models given by $\vec{X}$ at every infinitesimal time increment. We will compute the effect of such model averaging in the $n \to \infty$ limit next.

We note that $\sum_{i=1}^n |\dot{x}_i|^2$ is symmetric in all the x_is, so is the V. If the terminal value condition is symmetric in i also (for instance if $\rho_n(0)$ is all we observe and $f(\rho_n) := f(\vec{x})$ is symmetric in x_i), then we should consider

$$-\frac{\nu}{n}\log E\big[e^{n\nu^{-1}\{f(\rho_n(T))+\int_0^T V(\rho_n(t))\,dt\}}\big].$$

In the limit, by law of large number, $\{\rho_n(\cdot)\colon\ n = 1, 2, \ldots\}$ converges in probability to solution to the heat equation $\partial_t \rho = \nu \Delta \rho$. By Laplace principle,

$$\lim_{n\to\infty} -\frac{\nu}{n} \log E\big[e^{n\nu^{-1}\{f(\rho_n(T))+\int_0^T V(\rho_n(t))\,dt\}}\big]$$
$$= \inf\left\{ -f\big(\rho(T)\big) - \int_0^T V\big(\rho(s)\big)\,ds + K_T\big[\rho(\cdot)\big]\colon\ \rho(0) = \rho_0,\ \rho(T) = \rho_1,\ \rho(\cdot) \in C\big([0, T]; \mathcal{P}\big(\mathbb{R}^d\big)\big) \right\},$$

where the "kinetic energy" (make analogy with the finite dimensional situation)

$$K_T\big[\rho(\cdot)\big] := \lim_{\epsilon\to 0+} \lim_{n\to\infty} -\frac{\nu}{n} \log P\big(\rho_n(\cdot) \in B_\epsilon\big(\rho(\cdot)\big)\big).$$

One would expect

$$K_T\big[\rho(\cdot)\big] = \frac{1}{2} \int_0^T \big\|\dot{\rho}(t)\big\|^2_{-1,\rho(t)}\,dt$$

which is the continuum particle limit of finite dimensional situation

$$\frac{1}{2} \int_0^t \frac{1}{n} \sum_{i=1}^n \big|\dot{x}_i(t)\big|^2\,dt.$$

The interesting thing is, this is false. The correct answer is

$$K_T\big[\rho(\cdot)\big] = \frac{1}{2} \int_0^T \big\|\dot{\rho}(t) - \nu \Delta \rho\big\|^2_{-1,\rho(t)}\,dt, \tag{A.11}$$

which is the kinetic energy we introduced in this article. The rigorous derivation of (A.11) is a type of large deviation result (see for instance [10]). See also Chapter 13 of [14] for an approach linking with Hamilton–Jacobi equations in the space of measures.

To explain this in a nutshell, we note that for each $\rho_n(t)$ with t fixed, there exists $n!$ ways of permuting the ordered tuplet $(X_1(t), \ldots, X_n(t))$ and still obtain the same $\rho_n(t)$. In other words, the law of ρ_n keep averaging the $n!$ possible laws of the tuplet at a continuous time level. Therefore, the motion of ρ_n in an infinitesimal time is the motion of the tuplet plus the permutation combined. In the $n \to \infty$ limit, such permutation effect introduces an entropy dissipation mechanism for the continuum system.

We denote $C_{\rho_0,\rho_1}([0, T]; \mathcal{P}(\mathbb{R}^d))$ the collection of all $\mathcal{P}(\mathbb{R}^d)$-valued path $\rho(\cdot)$ with $\rho(0) = \rho_0$ and $\rho(T) = \rho_1$ and with continuous trajectory. Suppose for now that we can construct a $\mathcal{P}(\mathbb{R}^d)$-valued Brownian bridge process $\rho(\cdot)$ with probability law $P \in \mathcal{P}(C_{\rho_0,\rho_1}([0, T]; \mathcal{P}(\mathbb{R}^d)))$ (a rigorous theory for such process does not seem to exist in probability literature yet). Let P^V be the renormalized version of P taking into account of potential function V:

$$\frac{dP^V}{dP} = Z_V^{-1} e^{\int_0^T V(\rho(s))\,ds}.$$

Then, by analogy with the finite dimensional case,

$$\log E^P\big[e^{\int_0^T V(\rho(s))\,ds}\big] = -\inf\big\{R\big(Q \| P^V\big)\colon\ Q \in \mathcal{P}\big(C_{\rho_0,\rho_1}\big([0, T]; \mathcal{P}\big(\mathbb{R}^d\big)\big)\big)\big\}.$$

Again, R is strictly convex in Q with compact level set in weak convergence (narrow) topology. Therefore, the variational problem has a unique minimizer. We conjecture that such minimizer should be given by solution to a stochastic partial differential equation of the Euler type. In the "small noise" limit (corresponds to the $n \to \infty$ limit when considering ρ_n), it is expected to "converge" to Eqs. (1.1). Given the above discussion, it would be interesting to study well-posedness for such a stochastic Euler equation in the uniqueness of corresponding probability measure sense (that is, weak uniqueness).

Similar to the finite dimensional case, the Hamilton–Jacobi PDE theory extends now to infinite dimensional state space $\mathcal{P}(\mathbb{R}^d)$. Well-posedness of such equation is studied in this article. At least formally, (A.9) becomes

$$\dot{\rho} = \operatorname{grad} U(t, \rho). \tag{A.12}$$

Appendix B. Two technical lemmas

Lemma B.1. *Let S be a metric space and $f, g : S \mapsto \mathbb{R} \cup \{\pm\infty\}$ be measurable functions. Suppose that*

(1) *g is lower semicontinuous and bounded from below,*
(2) *$|f(x)| \leqslant \zeta(g(x))$ with an increasing sub-linear function $\zeta : \mathbb{R} \mapsto \mathbb{R}$,*
(3) *f is continuous on finite level sets of g, that is,*

$$\lim_{n\to\infty} f(x_n) = f(x), \quad \text{whenever } x_n \to x \text{ and } \sup_n g(x_n) < \infty.$$

Then $f + g$ is lower semicontinuous and bounded from below, and $f - g$ is upper semicontinuous and bounded from above.

Proof. We prove the case for $f + g$ only since the other one is similar.

The bound from below property follows because

$$C + \frac{1}{2} g \leqslant -\zeta(g) + g \leqslant f + g,$$

for some constant $C \in \mathbb{R}$. The above also implies that each finite level set of $f + g$ is contained in another finite level set of g. By continuity of f on finite level set of g, $f + g$ is lower semicontinuous. □

Lemma B.2. *Let S be a complete separable metric space with Borel σ-field $\mathcal{F}$. $\mu_n, \mu \in \mathcal{P}(S)$ are probability measures on S. Let $f, g : S \mapsto \mathbb{R} \cup \{\pm\infty\}$ be measurable functions on S satisfying the hypothesis of Lemma* B.1. *Further suppose that*

(1) *$\mu_n \Rightarrow \mu$ in the weak convergence of probability measure topology,*
(2) $\sup_n \int_S g \, d\mu_n < \infty$.

Then

$$\lim_{n\to\infty} \int_S f \, d\mu_n = \int_S f \, d\mu.$$

Proof. By Skorohod representation (e.g. Theorem 1.8 on p. 102 of [12]), there exists a reference probability space and S-valued random variables X_n, X such that $X_n \to X$ almost surely and $P(X_n \in dx) = \mu_n(dx)$, $P(X \in dx) = \mu(dx)$. Therefore,

$$\int_S f \, d\mu_n = E\big[f(X_n)\big], \qquad \int_S f \, d\mu = E\big[f(X)\big], \qquad \int_S g \, d\mu_n = E\big[g(X_n)\big].$$

From $\sup_n E[g(X_n)] < \infty$, by lower semicontinuity of g and by Fatou's lemma, we have $E[g(X)] < \infty$.

Let $\epsilon > 0$. By Lemma B.1, $f + \epsilon g$ is lower semicontinuous and bounded from below on S and $f - \epsilon g$ is upper semicontinuous and bounded from above. By Fatou's lemma,

$$\liminf_{n\to\infty} E\big[f(X_n) + \epsilon g(X_n)\big] \geqslant E\big[f(X)\big] + \epsilon E\big[g(X)\big],$$
$$\limsup_{n\to\infty} E\big[f(X_n) - \epsilon g(X_n)\big] \leqslant E\big[f(X)\big] - \epsilon E\big[g(X)\big].$$

Taking $\epsilon \to \infty$, $\lim_{n\to\infty} E[f(X_n)] = E[f(X)]$, we conclude the proof. □

References

[1] Robert A. Adams, John J.F. Fournier, Sobolev Spaces, second ed., Pure and Applied Mathematics Series, vol. 140, Academic Press, 2003.

[2] Luigi Ambrosio, Nicola Gigli, Giuseppe Savaré, Gradient Flows in Metric Spaces and the Wasserstein Spaces of Probability Measures, Lectures in Mathematics. ETH Zurich, Birkhäuser, 2005.

[3] Jean-David Benamou, Yann Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numer. Math. 84 (2000) 375–393.

[4] Yann Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (4) (1991) 375–417.

[5] Dario Cordero-Erausquin, Wilfrid Gangbo, Christian Houdré, Inequalities for generalized entropy and optimal transportation, in: Recent Advances in the Theory and Applications of Mass Transport, in: Contemp. Math., vol. 353, Amer. Math. Soc., Providence, RI, 2004, pp. 73–94.

[6] Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part I, J. Funct. Anal. 62 (3) (1985) 379–396; Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part II, J. Funct. Anal. 65 (3) (1986) 368–405; Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part III, J. Funct. Anal. 68 (2) (1986) 214–247; Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part IV, J. Funct. Anal. 90 (2) (1990) 237–283; Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part V, J. Funct. Anal. 97 (2) (1991) 417–465; Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part VII, J. Funct. Anal. 125 (1) (1994) 111–148.

[7] Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions, Part VI: Nonlinear A and Tataru's method refined, in: Evolution Equations, Control Theory, and Biomathematics, Han sur Lesse, 1991, in: Lecture Notes in Pure and Appl. Math., vol. 155, Dekker, New York, 1994, pp. 51–89.

[8] Michael G. Crandall, Hitoshi Ishii, Pierre-Louis Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. 27 (1992) 1–67.

[9] Amir Dembo, Ofer Zeitouni, Large Deviations, Techniques and Applications, second ed., Applications of Mathematics, vol. 38, Springer, New York, 1998.

[10] Donald A. Dawson, Jurgen Gärtner, Large deviations from the McKean–Vlasov limit for weakly interacting diffusions, Stochastics 20 (4) (1987) 247–308.

[11] Paul Dupuis, Richard Ellis, A Weak Convergence Approach to the Theory of Large Deviations, John Wiley & Sons, Inc., New York, 1997.

[12] Stewart Ethier, Thomas G. Kurtz, Markov Processes, John Wiley & Sons, Inc., New York, 1986.

[13] Jin Feng, Markos Katsoulakis, A Comparison Principle for Hamilton–Jacobi equations related to controlled gradient flows in infinite dimensions, Arch. Ration. Mech. Anal. 192 (2) (2009) 275–310.

[14] Jin Feng, Thomas G. Kurtz, Large Deviations for Stochastic Processes, American Mathematical Society, Providence, 2006.

[15] Wilfrid Gangbo, Truyen Nguyen, Adrian Tudorascu, Euler–Poisson systems as action-minimizing paths in the Wasserstein space, Arch. Ration. Mech. Anal. 192 (3) (2009) 419–452.

[16] Wilfrid Gangbo, Truyen Nguyen, Adrian Tudorascu, Hamilton–Jacobi equations in the Wasserstein space, Methods Appl. Anal. 15 (2) (2008) 155–183.

[17] Ugo Gianazza, Giuseppe Savaré, Giuseppe Toscani, The Wasserstein gradient flow of the Fisher information and the quantum drift–diffusion equation, Arch. Ration. Mech. Anal. 194 (1) (2009) 133–220.

[18] Richard Jordan, David Kinderlehrer, Felix Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.

[19] Thomas G. Kurtz, Richard H. Stockbridge, Stationary solutions and forward equations for controlled and singular martingale problems, Electron. J. Probab. 6 (15) (2001) 1–52.

[20] Cornelius Lanczos, The Variational Principles of Mechanics, fourth ed., Dover Publications, New York, 1970.

[21] Jean-Michel Lasry, Pierre-Louis Lions, Mean field games, Jpn. J. Math. 2 (2007) 229–260.

[22] Felix Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (1–2) (2001) 101–174.

[23] Lawrence Perko, Differential Equations and Dynamic Systems, Texts in Applied Mathematics, Springer, 2001.

[24] Cedric Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003.

[25] Cedric Villani, Optimal Transport: Old and New, Fundamental Principles of Mathematical Sciences, vol. 338, Springer, 2009.